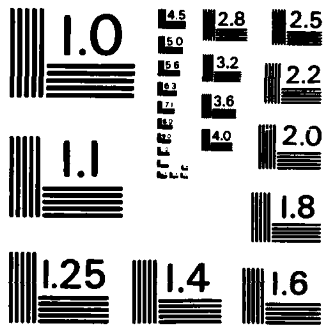
AD-A161 749 STATIONARY TIME SERIES QUANTILE FUNCTIONS NONPARAMETRIC INFERENCE AND RAN..(U) TEXAS A AND M UNIV COLLEGE STATION DEPT OF STATISTICS A HARPAZ SEP 85 TR-A-31 ARO-20140 N00014-84-K-0350 1/1

UNCLASSIFIED F/G 12/1 NL

END
FILMED
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

ARO 20140.7-MA

TEXAS A&M UNIVERSITY
COLLEGE STATION, TEXAS 77843-3143

Department of
STATISTICS
Phone 409 845-3141

STATIONARY TIME SERIES, QUANTILE FUNCTIONS, NONPARAMETRIC INFERENCE AND RANK TRANSFORM SPECTRUM

by Avraham Harpaz
Department of Statistics
Texas A&M University

Technical Report No. A-31

September 1985

Texas A&M Research Foundation
Project Nos. 5104 and 4858

"Multiple Time Series Modeling and Time Series Theoretic Statistical Methods"

and

"Functional Statistical Data Analysis and Modeling"

Sponsored by the Army Research Office and Office of Naval Research

Contracts N00014-84-K-0350 and DAAG29-83-K-0051

Professor Emanuel Parzen, Principal Investigator

Approved for public release; distribution unlimited

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER ARO 20140.7-MA	2. GOVT ACCESSION NO. AD-A161749	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Stationary Time Series, Quantile Functions, Nonparametric Inference and Rank Transform Spectrum		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Avraham Harpaz		8. CONTRACT OR GRANT NUMBER(s) N00014-84-K-0350 DAAG29-83-K-0051
9. PERFORMING ORGANIZATION NAME AND ADDRESS Texas A&M University Institute of Statistics College Station, TX 77843		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE September 1985
		13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Stationary Time Series, Quantile Functions, Nonparametric Tests, Rank Transform Spectrum, Wilcoxon Tests, Linear Rank Statistics.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In this dissertation, weak convergence results for dependent sequences are used to derive the asymptotic distribution of linear rank statistics for the two sample problem. It is shown that the asymptotic variance of linear rank statistics when computed from two independent time series depends on the spectrum of the rank transform time series. The behavior of the rank transform spectrum in terms of its relations to the original spectrum is also empirically examined.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ABSTRACT

Stationary Time Series, Quantile Functions, Nonparametric Inference and Rank Transform Spectrum. (December 1985)

Avraham Harpaz, B.S., AGEC Hebrew U. of Jerusalem;

M.S., AGEC Hebrew U. of Jerusalem

Chairman of Advisory Committee: Dr. Emanuel Parzen

In this dissertation, weak convergence results for dependent sequences are used to derive the asymptotic distribution of linear rank statistics for the two sample problem. It is shown that the asymptotic variance of linear rank statistics when computed from two independent time series depends on the spectrum of the rank transform time series. The behavior of the rank transform spectrum in terms of its relations to the original spectrum is also empirically examined.

TABLE OF CONTENTS

Table of Contents (Continued)

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 The Problem

Statistical theory can be divided into two categories: independence and dependence. In the independence category, the basic setting is a sequence of independent random variables (r.v.s) (possibly multivariate). This category contains the areas of parametric and nonparametric inference and estimation, regression, analysis of variance, multivariate analysis and more. In the second category, the basic setting is a sequence of dependent r.v.s. This category contains the areas of time series analysis and stochastic processes (Markov chains, renewal theory, etc.).

There is a growing literature on the extension of asymptotic theorems from the independence case to the dependence case by introducing concepts of asymptotic independence. Unfortunately, most of the published results have not been expressed in forms that are easy to apply in statistical procedures. In this work our goal is to interpret and extend the theory of non-parametric tests for time series to make it usable in practical data analysis procedures. In particular, the asymptotic theory of linear rank statistics in the two sample problem is extended to express the asymptotic variance in an interpretable and estimable form. Another part of this work is an

---

This dissertation will follow the format for the *Journal of the American Statistical Association*.

empirical examination of the properties of the spectrum of rank transform time series as compared to the spectrum of the original time series.

1.2 Literature Review

The empirical process and the quantile process serve as the main ingredients in the theory and procedures considered in this work. The theory of the weak convergence of these processes to Gaussian processes, in the independent case, is covered in great detail in Billingsley (1968) and in Csörgő and Révész (1981). For the two sample problem, Pyke and Shorack (1968) introduced a two sample empirical process which can be used to represent linear rank statistics, and developed the asymptotic distribution of their process for the independence case. The asymptotic theory in the dependence case of stochastic processes which can be used to represent linear rank statistics is an active area of research. Billingsley (1968) gave the basic weak convergence result for the empirical process under ϕ-mixing conditions. Sen (1971) has improved Billingsley's result by weakening the required mixing conditions. Yoshihara (1978) and Yokoyama (1980) have successively improved this result by weakening twice more the required mixing conditions. The weak convergence of the empirical process has been established also for two other notions of mixing (see, for example, Yokoyama (1973) and Mehra and Rao (1975b) for results under strong mixing conditions and Gastwirth and Rubin (1975) for introduction of strong mixing Δ_s conditions and weak convergence results under those conditions). The

weak convergence of the quantile process is usually established via a Bahadur representation type result (which is stronger than just implying the weak convergence of the quantile process). Babu and Singh (1978) give such a result for both ϕ-mixing and strong mixing sequences.

An application of the weak convergence of the empirical process to the two sample problem is given in Fears and Mehra (1974) for ϕ-mixing sequences. They extend Pyke and Shorack (1968) to include the dependence case. Mehra and Rao (1975a) applied the weak convergence of the empirical process to obtain asymptotic results for functions of order statistics for both ϕ-mixing and strong mixing sequences. Another application of that type is given in Falk and Kohne (1984). They study the behavior of the sign test under mixing conditions and suggest a way to adjust the critical region to the dependence case.

None of the above papers take the approach considered in this work, which is to express the asymptotic variance of the statistics studied in terms of the spectral density of some related time series, which we call the rank transform time series.

CHAPTER II

ASYMPTOTIC THEORY FOR EMPIRICAL AND QUANTILE PROCESSES

2.1 Distribution Functions and Quantile Functions - Definitions and Basic Properties

Let X be a random variable (RV) defined on a probability space (Ω, A, P). Then its *distribution function* (DF), F, is defined by

$$F(x) = \Pr(X \le x) = P\{\omega \in \Omega : X(\omega) \le x\}, \quad -\infty < x < \infty \ . \qquad (2.1.1)$$

The *quantile function* (QF), Q, corresponding to F is defined by

$$Q(u) = F^{-1}(u) = \inf\{x \in \mathbb{R} : F(x) \ge u\} \ , \quad 0 \le u \le 1 \ . \qquad (2.1.2)$$

Note that $Q(0)$ is formally $-\infty$ and when $F(x)<1$ for all $x \in \mathbb{R}$ then $Q(1)$ is formally $+\infty$.

When F is continuous and strictly increasing then Q is its true inverse. In general, however, we have

$$F(x) = \sup\{u \in [0,1] : Q(u) \le x\}, \quad -\infty < x < \infty \ . \qquad (2.1.3)$$

From these definitions it follows that F is right-continuous, Q is left-continuous and both F and Q are nondecreasing. Some of the basic relationships between F and Q are summarized in the following theorem.

Theorem 2.1.1 : Let F be a DF with corresponding quantile function Q. Then

a) $F(Q(u)) \geq u, \quad 0 \leq u \leq 1$, (2.1.4)

with equality if F is continuous at Q(u).

b) $F(x) \geq u$ iff $Q(u) \leq x$. (2.1.5)

Proof: The proof of these standard facts is given here in order to illustrate the methods by which one can relate the properties of distribution functions and quantile functions.

a) First observe that for $u=0$, 2.1.4 is satisfied trivially since F is nonnegative. Also, when $Q(u)=+\infty$, then $F(Q(u))=F(+\infty)=1\geq u$ since $0\leq u\leq 1$. Next, for $0<u\leq 1$ and when $Q(u)<+\infty$, it follows from the definition of Q that

$$F(Q(u)-\epsilon) < u \leq F(Q(u)+\epsilon), \quad \text{any } \epsilon>0 .$$

Hence, letting $\epsilon \longrightarrow 0$ and using the right-continuity of F, we have

$$F(Q(u)) \geq u ,$$

and if F is continuous at Q(u) we also have

$$F(Q(u)) \leq u ,$$

which gives the equality.

b) Assume that $Q(u) \leq x$. Then the monotonicity of F and part a) above give

$$F(x) \geq F(Q(u)) \geq u .$$

Now assume that $F(x) \geq u$. Then

$$x \in \{z : F(z) \geq u\}$$

and hence, using the definition of Q,

$$Q(u) \leq x \ .$$

Q.E.D

Let I_A be the indicator function of a set A defined by

$$I_A(t) = \begin{cases} 1 \text{ if } t \in A \\ 0 \text{ if } t \notin A \ . \end{cases} \tag{2.1.6}$$

Let X(1),X(2),...,X(n) be RVs defined on a common probability space (Ω,A,P) and having common DF F and QF Q. The *sample DF* F_n (also called the *empirical DF* and denoted *EDF*) is defined by

$$F_n(x) = (1/n) \sum_{i=1}^{n} I_{(-\infty, x]}(X(i)) \ , \quad -\infty<x<\infty \ . \tag{2.1.7}$$

The corresponding *sample QF* Q_n is defined by

$$Q_n(u) = \inf\{x : F_n(x) \geq u\} \ , \quad 0<u\leq 1 \ . \tag{2.1.8}$$

In terms of X(1:n),X(2:n),...,X(n:n), the order statistics corresponding to X(1),X(2),...,X(n), a formula for Q_n is :

$$Q_n(u) = X(j{:}n) \quad \text{for} \quad (j-1)/n < u \leq j/n \ , \ j=1,\ldots,n \ . \tag{2.1.9}$$

Note that $Q_n(0)$ was left undefined by the above definition. Parzen (1979) suggests that $Q_n(0)$ be taken to be a natural minimum when one is known (e.g. for nonnegative RV one can take $Q_n(0) = 0$).

2.2 Asymptotic Distribution of F_n and Q_n : Independent Case

In this section we assume that the RVs $X(1),\ldots,X(n)$ are independent and present some of the basic properties of F_n and Q_n as estimators of F and Q respectively. The case of dependent RVs (i.e. time series) will be treated below.

For fixed $x \in \mathbb{R}$, $F_n(x)$ is the average of independent and identically distributed (IID) RVs each distributed as $\mathcal{B}(1,F(x))$ ($\mathcal{B}$inomial distribution with parameters 1 and F(x)). Therefore,

$$nF_n(x) \sim B(n,F(x)) \ ,$$

$$E[F_n(x)] = F(x) \ ,$$

$$Var[F_n(x)] = (1/n)F(x)(1-F(x)) \quad .$$

Using the properties of the binomial distribution, the strong law of large numbers (SLLN) and the IID version of the central limit theorem (CLT), we immediately have the following result.

Theorem 2.2.1 : Consistency and Asymptotic Normality of the Sample Distribution Function. For any fixed $x \in \mathbb{R}$,

a) $F_n(x) \longrightarrow F(x)$ (2.2.1)

where the convergence holds in probability (denoted $\xrightarrow{p}$), in quadratic mean (denoted $\xrightarrow{qm}$) and almost surely (denoted $\xrightarrow{as}$).

b) $n^{1/2}[F_n(x)-F(x)] \xrightarrow{d} N(0,F(x)(1-F(x)))$. (2.2.2)

For Q_n we have the following results.

Theorem 2.2.2 : Consistency of the Sample Quantile Function. Let $0<u<1$ and assume that Q is continuous at u. Then

$$Q_n(u) \xrightarrow{as} Q(u) \ . \tag{2.2.3}$$

Proof: Let $\epsilon>0$. Then by the definition of Q and the assumed continuity of Q at u, we have

$$F(Q(u)-\epsilon) < u < F(Q(u)+\epsilon) \ .$$

From Theorem 2.2.1 (a) it follows that

$$F_n(Q(u)-\epsilon) \xrightarrow{as} F(Q(u)-\epsilon)$$

and

$$F_n(Q(u)+\epsilon) \xrightarrow{as} F(Q(u)+\epsilon) \ .$$

Hence,

$$\Pr[F_m(Q(u)-\epsilon) < u < F_m(Q(u)+\epsilon) \ , \text{ all } m\geq n] \longrightarrow 1 \quad \text{as} \quad n\to\infty \ .$$

Thus, by Theorem 2.1.1 (b),

$$\Pr[(Q(u)-\epsilon) < Q_m(u) \leq (Q(u)+\epsilon) \ , \text{ all } m\geq n] \longrightarrow 1 \quad \text{as} \quad n\to\infty \ ,$$

which is equivalent to

$$\Pr[\sup_{m\geq n}|Q_m(u)-Q(u)| \leq\epsilon] \longrightarrow 1 \quad \text{as} \quad n\to\infty \ ,$$

giving the result.

Q.E.D

The asymptotic normality of $Q_n(u)$ (at fixed $0<u<1$) can be established under the assumption that F possesses a density f in a neighborhood of Q(u) and that f is positive and continuous at Q(u) (see Serfling (1980), Theorem A p. 77). An alternative approach (adopted in this work) is to assume that the density f is differentiable at Q(u) and use the Bahadur representation. This important result, first given by Bahadur (1966), is stated in the following theorem (for a proof, see Serfling (1980), pp. 91-95).

Theorem 2.2.3 : Bahadur Representation. Let $0<u<1$. Assume that F is twice differentiable at Q(u), with

$$fQ(u) \equiv f(Q(u)) = F'(Q(u)) > 0 .$$

Then

$$Q_n(u) = Q(u) - [F_n(Q(u))-u]/fQ(u) + R_n , \tag{2.2.4}$$

where with probability one

$$R_n = O(n^{-3/4}(\log(n))^{3/4}) , \quad \text{as } n\to\infty . \tag{2.2.5}$$

The asymptotic normality of $Q_n(u)$ can be proved as an immediate consequence of the Bahadur representation and the asymptotic normality of F_n.

Theorem 2.2.4 : Asymptotic Normality of the Sample Quantile Function. Assume that f is differentiable at Q(u) with $fQ(u)>0$. Then

$$n^{1/2}[Q_n(u)-Q(u)] \xrightarrow{d} N(0,\sigma_u^2) \tag{2.2.6}$$

where

$$\sigma_u^2 = u(1-u)/[fQ(u)]^2 \ . \qquad (2.2.7)$$

Proof: By Theorem 2.2.3, we have

$$n^{1/2}[Q_n(u)-Q(u)] = -n^{1/2}[F_n(Q(u))-u]/fQ(u) + n^{1/2}R_n$$

and

$$n^{1/2}R_n \xrightarrow{p} 0 \ .$$

The result now follows using Theorem 2.2.1 (b), Slutsky's Theorem and the fact that

$$\text{if} \quad Z \sim N(0,\sigma^2) \quad \text{then} \quad -Z/a \sim N(0,\sigma^2/a^2) \ .$$

Q.E.D.

2.3 Empirical and Quantile Processes - Independent Case

The behavior of F_n and Q_n when regarded as random functions or stochastic processes is the theoretical basis of modern approaches to deriving the distribution theory for many statistical techniques. Define the *empirical process* $\{\alpha_n(u), 0 \leq u \leq 1\}$ by

$$\alpha_n(u) = n^{1/2}[F_n(Q(u))-F(Q(u))] \ , \quad 0 \leq u \leq 1 \ , \qquad (2.3.1)$$

and the *quantile process* $\{\beta_n(u), 0 \leq u \leq 1\}$ by

$$\beta_n(u) = n^{1/2}[Q_n(u)-Q(u)] \ , \quad 0 \leq u \leq 1 \ . \qquad (2.3.2)$$

First we present an asymptotic normality type result for α_n. The ideas and techniques required for this type of result are those of

convergence of probability measures on metric spaces. A detailed and comprehensive treatment of this topic can be found in Billingsley (1968). In this work we use these concepts and results somewhat heuristically.

Lemma 2.3.1 : Let U(1),U(2),...,U(n) be IID RVs with common DF G and such that

$$0 \le U(i,\omega) \le 1 \ , \ \omega\epsilon\Omega, \quad i=1,\ldots,n \ .$$

Define the random element of D(0,1), α_n^U, by

$$\alpha_n^U(u) = n^{1/2}[G_n(u)-G(u)] \ , \quad 0\le u\le 1 \ ,$$

where G_n is the EDF corresponding to U(1),...,U(n). Then

$$\alpha_n^U \xrightarrow{d} \alpha^U \quad ,$$

where α^U is the Gaussian random element of D(0,1) specified by

$$E\{\alpha^U(u)\} = 0 \quad 0\le u\le 1 \quad ,$$

$$E\{\alpha^U(u)\cdot\alpha^U(v)\} = G(u)(1-G(v)) \ , \quad 0\le u\le v\le 1 \quad .$$

Proof: See Billingsley (1968) Theorem 16.4 p. 141 .

Remark: When U(1),...,U(n) are uniformly distributed on [0,1] (U(0,1)) then G(u)=u, 0≤u≤1 and α^U is distributed as the *Brownian bridge process*.

Theorem 2.3.1 : *Asymptotic Distribution of the Empirical Process.* Let X(1),...,X(n) be IID RVs with common DF F and let α_n be the

corresponding empirical process as defined by 2.3.1. Then

$$\alpha_n \xrightarrow{d} \alpha \quad ,$$

where α is the Gaussian random element of $D(0,1)$ specified by

$$E\{\alpha(u)\} = 0 \quad 0 \le u \le 1 \quad ,$$

$$E\{\alpha(u)\cdot\alpha(v)\} = F(Q(u))[1-F(Q(v))] \ , \quad 0 \le u \le v \le 1 \quad .$$

Proof: Define the RVs

$$U(i) = F(X(i)) \ , \quad i=1,\ldots,n \ .$$

Then $U(1),\ldots,U(n)$ are IID RVs with common DF

$$G(u) = F(Q(u)) \ , \quad 0 \le u \le 1 \ .$$

The EDF corresponding to $U(1),\ldots,U(n)$ is given by

$$G_n(u) = F_n(Q(u)) \ , \quad 0 \le u \le 1 \ .$$

Lemma 2.3.1 now implies

$$\alpha_n^U \xrightarrow{d} \alpha^U \quad ,$$

where α_n^U and α^U are defined as in Lemma 2.3.1 but in terms of G and G_n defined here. But

$$\alpha_n = \alpha_n^U \ \text{ and } \ \alpha \overset{d}{=} \alpha^U \quad ,$$

giving the result.

Q.E.D.

Note that when F is continuous then α is distributed as a Brownian bridge process.

The asymptotic distribution of β_n is derived from a result that can be regarded as an extention of the Bahadur representation as given in Theorem 2.2.3. This result was first established by Kiefer (1970) and then extended to the general case by Csörgő and Révész (1981). We list here this later result without proof.

Theorem 2.3.2 : Let X(1),...,X(n) be IID RVs with a common DF F which is twice differentiable on (a,b) where

$$-\infty \le a = \sup\{x : F(x)=0\} \quad \text{and} \quad +\infty \ge b = \inf\{x : F(x)=1\}$$

and

$$F' = f > 0 \quad \text{on } (a,b) .$$

Assume that F also satisfies

$$\sup_{a<x<b} \{F(x)(1-F(x))\,|f'(x)/f^2(x)|\} \le \gamma \quad \text{for some } \gamma>0 , \tag{2.3.3}$$

and

f is nondecreasing (nonincreasing) on an interval to the right of a (to the left of b) . (2.3.4)

Let

$$R_n = \sup_{0\le u\le 1} |(F_n(Q(u))-u)-(Q(u)-Q_n(u))fQ(u)| . \tag{2.3.5}$$

Then

$$\limsup_{n\to\infty} [n^{-3/4}(\log(n))^{1/2}(\log\log(n))^{1/4}]^{-1} R_n \stackrel{a.s}{=} 2^{-1/4} \quad . \tag{2.3.6}$$

Proof: See Csörgő and Révész (1981) Theorems 4.5.6 p. 149, 5.2.1 and 5.2.2 p. 160.

Remark: This result goes much beyond our needs here. We will extract from it the following two results about quantile processes (which are only part of the statement of the theorem).

Theorem 2.3.3 : Asymptotic Distribution of the Quantile Process. Under the conditions of Theorem 2.3.2, we have

a) $$fQ(\cdot)\beta_n(\cdot) \xrightarrow{d} \beta(\cdot) \ , \tag{2.3.7}$$

where β is distributed as a Brownian bridge process.

b) $$\beta \stackrel{a.s}{=} -\alpha \quad , \tag{2.3.8}$$

where $\stackrel{a.s}{=}$ means that there exists $\Omega_1 \subset \Omega$, with $P(\Omega_1)=1$ and such that for any $\omega \epsilon \Omega_1$ $\beta(u,\omega)=-\alpha(u,\omega)$ $0 \le u \le 1$.

Proof: Immediate from the convergence of the empirical process (Theorem 2.3.1) the extended Bahadur representation (Theorem 2.3.2) and Theorem 4.1 p. 25 of Billingsley (1968).

2.4 Empirical and Quantile Processes - Dependent Case

In this section we assume that the RVs $X(1),\ldots,X(n)$ come from a stochastic process (SP) $\{X\}=\{X(i):i\epsilon Z\}$, where Z is the set of integers. We first define the notions of stationarity and asymptotic independence required, and then present some asymptotic results for the empirical and quantile processes.

Definitions:

a) The SP $\{X(i):i\epsilon Z\}$ is said to be (strictly) stationary if for any $k\geq 0$, $i_1\leq i_2\leq \ldots \leq i_k$ and any $j\geq 0$, the vectors $(X(i_1),\ldots,X(i_k))$ and $(X(i_1+j),\ldots,X(i_k+j))$ are identically distributed.

b) For $I\subset Z$, we denote by B_I the σ-field generated by the RVs $\{X(i):i\epsilon I\}$. We now introduce the notion of *mixing*. The SP $\{X(i):i\epsilon Z\}$ is said to be ϕ-mixing if there exists a positive integer M and a function ϕ for which $\phi(m)\longrightarrow 0$ as $m\longrightarrow\infty$ such that

$$|P(A\cap B)-P(A)P(B)| \leq \phi(m)P(A) \tag{2.4.1}$$

whenever $m\geq M$, $n\epsilon Z$, $A\epsilon B_I$ and $B\epsilon B_J$ where $I=\{i:i\leq n\}$ and $J=\{j:j\geq m+n\}$.

Note that when $P(A)>0$ then 2.4.1 is equivalent to

$$|P(B|A)-P(B)| \leq \phi(m) \tag{2.4.2}$$

while when $P(A)=0$ then 2.4.1 is satisfied trivially. Hence taking the left side of 2.4.2 as 0 when $P(A)=0$ one can define the mixing function $\phi(m)$ by

$$\phi(m) = \sup\{|P(B|A)-P(B)|:A\epsilon B_I \text{ , } B\epsilon B_J\} \text{ , } \quad m\geq 0 \tag{2.4.3}$$

where I,J and n are as in 2.4.1. Then one places assumptions on ϕ in order to obtain asymptotic results.

Remark: Other related notions of mixing have been introduced in the literature. Two of them use the same l.h.s (left hand side) of 2.4.1 but put other bounds on the r.h.s (right hand side). One is the weak ϕ-mixing, where the r.h.s of 2.4.1 is replaced by $\phi(m)$ alone and the

other one is the strong ϕ-mixing in which the r.h.s of 2.4.1 is replaced by $\phi(m)P(A)P(B)$. Another measure of dependence is the one introduced by Gastwirth and Rubin (1975). Their definitions and results are particularly appropriate for time series which might not be ϕ-mixing. However, their conditions for the weak convergence of the empirical process do not cover all the time series models we would like to consider. We present here the results for processes satisfying only ϕ-mixing conditions. We do not discuss further these conditions because our approach is only to illustrate the kinds of conditions required to prove asymptotic normality in the dependent case. These results are used in our statistical analysis procedures only heuristically since we do not verify that the mixing conditions assumed in this section are satisfied by the time series that we observe.

The main result concerning the asymptotic behavior of the empirical process for ϕ-mixing RVs is given in Billingsley (1968). This result has been shown to hold under weaker conditions by Sen (1971). It has also been shown to hold under other definitions of mixing structures. We will list here Billingsley's result and mention Sen's improvement. This will fulfill the needs of our heuristic approach. Before listing the theorem, we first define the notion of *dependence DF*.

Definition: Let (X_1,X_2) be a bivariate RV with DF $F(\cdot,\cdot)$ and marginal DFs F_1, F_2 with corresponding QFs Q_1, Q_2. Then the *dependence DF* of (X_1,X_2) is defined by

$$B(u_1,u_2) = F(Q_1(u_1),Q_2(u_2)) \quad , \quad 0 \le u_1,u_2 \le 1 \quad . \tag{2.4.4}$$

Note that the dependence DF is, in fact, the DF of the bivariate RV (U_1,U_2) where $U_1=F_1(X_1)$ and $U_2=F_2(X_2)$.

Theorem 2.4.1 : (Billingsley (1968)) Let $\{X(i):i\epsilon Z\}$ be a stationary ϕ-mixing process with a continuous marginal DF F and such that

$$\sum_1^\infty n^2[\phi(n)]^{1/2} < \infty \quad . \tag{2.4.5}$$

Then

$$\alpha_n \xrightarrow{d} \alpha^F \tag{2.4.6}$$

where α_n is the empirical process corresponding to $\{X\}$ and α^F is the Gaussian random function specified by

$$E\{\alpha^F(u)\} = 0 \ , \quad 0 \le u \le 1 \ , \tag{2.4.7}$$

$$E\{\alpha^F(u)\alpha^F(v)\} = K_B(u,v) \ , \tag{2.4.8}$$

where $K_B(u,v)$ is the dependence distribution covariance kernel of the process $\{X(i):i\epsilon Z\}$, defined by

$$K_B(u,v) = u \wedge v - uv + 2\sum_1^\infty [B_k(u,v)-uv] \ , \ 0 \le u,v \le 1 \ , \tag{2.4.9}$$

in terms of B_k, the dependence DF of $(X(i),X(i+k))$.

Proof: See Billingsley (1968) Thecrem 22.1 p. 197 .

Remark: Sen (1971) showed that the same result holds under the weaker condition that

$$\sum_{1}^{\infty} n[\phi(n)]^{1/2} < \infty \quad . \qquad (2.4.10)$$

The asymptotic normality of the quantile process is derived in the same manner as in the independent case using a Bahadur representation type result for mixing processes. A result of that type is given in Babu and Singh (1978).

Theorem 2.4.2 : Bahadur Representation for Mixing Processes. Let $\{X(i):i\epsilon Z\}$ be a stationary ϕ-mixing process with $\sum_{1}^{\infty}[\phi(k)]^{1/2}<\infty$. Define the following conditions for the underlying DF F and its density f.

Condition 1: For some interval I, f' exists and is bounded on I, f vanishes outside I, $\inf\{f(x):x\epsilon I\}>0$ and $\sup\{f(x):x\epsilon I\}<\infty$.

Condition 2: For some $0\leq a\leq b\leq 1$ and $\epsilon>0$, condition 1 is satisfied for the interval $I=[Q(a)-\epsilon,Q(b)+\epsilon]$, except that f need not vanish outside that interval.

Let

$$c_n = n^{-3/4}[\log(n)]^{1/2}[\log\log(n)]^{1/4} \quad ,$$

$$R_n(u) = [F_n(Q(u))-u]-fQ(u)[Q(u)-Q_n(u)] \quad ,$$

$$R_n^{(1)} = \sup_{0\leq u\leq 1}|R_n(u)| \quad ,$$

$$R_n^{(2)} = \sup_{a\leq u\leq b}|R_n(u)| \quad .$$

Then, under condition 1,

$$\limsup_{n\to\infty} c_n^{-1}R_n^{(1)} \leq C_1 \quad \text{a.s} \quad , \qquad (2.4.11)$$

and, under condition 2,

$$\limsup_{n\to\infty} c_n^{-1} R_n^{(2)} \le C_2 \quad \text{a.s} \quad , \tag{2.4.12}$$

where C_1 and C_2 are positive constants.

Proof: See Babu and Singh (1978) Theorem 7 and Remark 4.2 . The asymptotic normality is now given by the following theorem.

Theorem 2.4.3 : Asymptotic Normality of the Quantile Process. Let $\{X(i):i\in Z\}$ be a stationary ϕ-mixing process with $\sum_1^\infty k^2[\phi(k)]^{1/2}<\infty$. Let β_n be the quantile process formed from a realization of length n from $\{X\}$ and let $\beta^F=-\alpha^F$ where α^F is the Gaussian process defined in Theorem 2.4.1. Then, under condition 1 of Theorem 2.4.2, we have

$$\{fQ(u)\beta_n(u):0\le u\le 1\} \xrightarrow{d} \{\beta^F(u):0\le u\le 1\} \quad , \tag{2.4.13}$$

while under condition 2 of Theorem 2.4.2, we have

$$\{fQ(u)\beta_n(u):a\le u\le b\} \xrightarrow{d} \{\beta^F(u):a\le u\le b\} \quad . \tag{2.4.14}$$

Proof: Immediate from the convergence of the empirical process (Theorem 2.4.1), the Bahadur representation (Theorem 2.4.2) and Theorem 4.1 p. 25 of Billingsley (1968).

2.5 Spectral Density Interpretation of $K_B(u,v)$

Let $\{X\}$ be a stationary time series with absolutely summable correlation function $\rho_X(k)$ defined by

$$\rho_X(k) = \text{Corr}[X(i),X(i+k)] \quad , \ i,k\in Z \quad . \tag{2.5.1}$$

Let $f(\omega;X)$ be the corresponding spectral density function defined by

$$f(\omega;X) = 1+2\sum_{1}^{\infty}\rho_X(k)\cos(2\pi k\omega) \quad , \; 0\le\omega\le1 \quad . \qquad (2.5.2)$$

Let $\bar{x}_n$ be the mean of a sample $X(1),\ldots,X(n)$ from $\{X\}$. Assume that the time series $\{X\}$ obeys the conditions for $\bar{x}_n$ to be asymptotically normal. Then

$$n^{1/2}[\bar{x}_n-\mu_X] \xrightarrow{d} N(0,\sigma^2) \qquad (2.5.3)$$

where μ_X is the common mean, $\mu_X=E[X(i)]$, and the asymptotic variance σ^2 can be expressed in terms of the value at zero frequency of the spectral density function $f(\omega;X)$:

$$\sigma^2 = \mathrm{Var}[X(1)]f(0;X) \quad . \qquad (2.5.4)$$

The formula for σ^2 can be derived directly by calculating the variance of $\bar{x}_n$ and taking the appropriate limit. We would like to present here a method for deriving results like the above which will be used later to derive the asymptotic variance of linear rank statistics.

Let g be a function such that the time series $\{g(X)\}$ possesses a spectral density function, and assume that the time series $\{X\}$ satisfies the conditions for the asymptotic normality of the empirical process α_n. Define

$$T_n(g) = (1/n)\sum_{i=1}^{n} g(X(i))$$
$$= \int_{R} g(x)dF_n(x) = \int_0^1 g(Q(u))dF_n(Q(u)) \;, \tag{2.5.5}$$

and

$$\Delta_n(g) = n^{1/2}[T_n(g)-\int_0^1 g(Q(u))du]$$
$$= \int_0^1 g(Q(u))d\alpha_n(u) \quad . \tag{2.5.6}$$

Then

$$\Delta_n(g) \xrightarrow{d} \Delta(g) = \int_0^1 g(Q(u))d\alpha(u) \tag{2.5.7}$$

and $\Delta(g)$ is a $N(0,\sigma_g^2)$ RV whose variance can be expressed as

$$\sigma_g^2 = \mathrm{Var}[g(X)]f(0;g(X)) \tag{2.5.8}$$

where $f(\cdot;g(X))$ is the spectral density of the time series $\{g(X)\}$. We derive 2.5.8 by a method to be used below for linear rank statistics:

$$\sigma_g^2 = \int_0^1\int_0^1 g(Q(u))g(Q(v))dK_B(u,v)$$
$$= \int_0^1 [g(Q(u))]^2du - \{\int_0^1 g(Q(u))du\}^2$$
$$+ 2\sum_1^\infty\int_0^1\int_0^1 g(Q(u))g(Q(v))[b_k(u,v)-1]dudv$$
$$= \mathrm{Var}[g(X)]\{1+2\sum_1^\infty \mathrm{Corr}[g(X(1)),g(X(1+k))]\}$$
$$= \mathrm{Var}[g(X)]f(0;g(X)) \tag{2.5.9}$$

where b_k is the dependence density function corresponding to the dependence DF B_k. Note that taking $g(x)=x$ in 2.5.9 gives the result 2.5.4 for $\bar{x}_n$ as a special case.

CHAPTER III

ASYMPTOTIC THEORY FOR TWO-SAMPLE LINEAR RANK STATISTICS

3.1 General Settings and Assumptions

We consider two samples $X(1),\ldots,X(m)$ and $Y(1),\ldots,Y(n)$ respectively representing finite realizations from two strictly stationary stochastic processes $\{X(i):i\epsilon Z\}$ and $\{Y(j):j\epsilon Z\}$, where $Z=\{0,\pm1,\pm2,\ldots\}$ is the set of all integers.

Assumption 1: The two processes are independent.

Assumption 2: The univariate marginal DFs of the two processes are respectively given by

$$F(x) = Pr[X(i)\leq x]\ ,\quad i\epsilon Z\ ,\quad x\epsilon\mathcal{R}$$

and

$$G(y) = Pr[Y(j)\leq y]\ ,\quad j\epsilon Z\ ,\quad y\epsilon\mathcal{R}\ . \tag{3.1.1}$$

We assume that F and G are continuous and denote their corresponding QFs by Q_F and Q_G respectively.

Remark: One often desires to test the null hypothesis $H_0:F(x)=G(x)$ all x.

Assumption 3: We further assume that each of the processes $\{X\}$ and $\{Y\}$ satisfies the conditions on the dependence structure and on the smoothness of the DFs, required for the convergence of the empirical and quantile processes to the appropriate Gaussian processes, as described in the preceding chapter.

To compare the univariate marginal distributions of the two time series, we will follow the development as in Parzen (1983) adjusting it to the case of time series. Additional assumptions will be made as needed along the way.

3.2 Definitions and Notations

As above, we let $X(1),\ldots,X(m)$ and $Y(1),\ldots,Y(n)$ be the observed samples and F, G, Q_F and Q_G be the marginal DFs and QFs respectively. We define the sample functions $\tilde{F}$, $\tilde{G}$, $\tilde{Q}_F$ and $\tilde{Q}_G$ by

$$\tilde{F}(x) = (1/m)\sum_{j=1}^{m} I_{(-\infty,x]}(X(j)) \quad -\infty<x<\infty , \tag{3.2.1}$$

$$\tilde{Q}_F(u) = \inf\{x:\tilde{F}(x)\geq u\} \quad 0\leq u\leq 1 , \tag{3.2.2}$$

$$\tilde{G}(x) = (1/n)\sum_{i=1}^{n} I_{(-\infty,x]}(Y(i)) \quad -\infty<x<\infty , \tag{3.2.3}$$

$$\tilde{Q}_G(u) = \inf\{x:\tilde{G}(x)\geq u\} \quad 0\leq u\leq 1 . \tag{3.2.4}$$

Next we let $N=m+n$, $\lambda_N=m/N$ and define

$$\tilde{H}(x) = \lambda_N\tilde{F}(x) + (1-\lambda_N)\tilde{G}(x) , \quad -\infty<x<\infty \tag{3.2.5}$$

to be the EDF of the pooled sample $X(1),\ldots,X(m),Y(1),\ldots,Y(n)$. We assume that $\lambda_N\longrightarrow\lambda$ as $N\longrightarrow\infty$, with $0<\lambda<1$, and define

$$H(x) = H_\lambda(x) = \lambda F(x)+(1-\lambda)G(x) , \quad -\infty<x<\infty . \tag{3.2.6}$$

In general, we define the inverse of a nondecreasing, right or left continuous function, $D(t)$, by

$$D^{-1}(u) = \inf\{t:D(t)\geq u\} \tag{3.2.7}$$

if D is right-continuous, and by

$$D^{-1}(u) = \sup\{t : D(t) \le u\} \qquad (3.2.8)$$

if D is left-continuous.

Next we define the basic comparison functions

$$\tilde{D}_1(u) = \tilde{H}\tilde{Q}_F(u) = \tilde{H}[\tilde{F}^{-1}(u)] \ , \qquad (3.2.9)$$

as estimator of $D_1(u)=HQ_F(u)$, and

$$\tilde{D}(u) = \tilde{D}_1^{-1}(u) \ , \qquad (3.2.10)$$

as estimator of $D(u)=D_1^{-1}(u)=FQ_H(u)$. Observe that $\tilde{D}_1$ is left-continuous, so that, $\tilde{D}$ is right-continuous. Observe also that the Pyke-Shorack (1968) sample function, $\tilde{F}\tilde{Q}_H(u)$, which is another estimator of D(u), is not, in general, equal to $\tilde{D}(u)$. This can be seen from the following explicit formulas for $\tilde{D}(u)$ and $\tilde{F}\tilde{Q}_H(u)$ given in Parzen (1983):

$$\begin{aligned} \tilde{D}(u) &= 0 \ , \quad 0 \le u < (R_1/N) \ , \\ &= j/m \ , \quad (R_j/N) \le u < (R_{j+1}/N) \ , \quad j=1,\ldots,m-1 \ , \qquad (3.2.11) \\ &= 1 \ , \quad (R_m/N) \le u < 1 \ ; \end{aligned}$$

and

$$\tilde{F}\tilde{Q}_H(u) = 0 \quad , \ 0 < u \le (R_1-1)/N \quad ,$$

$$= j/m \quad , \ (R_j-1)/N < u \le (R_{j+1}-1)/N \ , \ j=1,\ldots,m-1 \ , \qquad (3.2.12)$$

$$= 1 \quad , \ (R_m-1)/N < u \le 1 \quad .$$

where R_j or $R(j)$, for $j=1,\ldots,m$, is defined to be the rank in the pooled sample of $X(j:m)$, the j^{th} order statistic in the X-sample. More precisely,

$$R_j \equiv R(j) = N\tilde{H}(X(j:m)) \ , \quad j=1,\ldots,m \quad . \qquad (3.2.13)$$

The null hypothesis $H_0:F\equiv G$ is often tested using a statistic of the form

$$T_N(J) = (1/m)\sum_{j=1}^{m} J(R(j)/(N+1)) \qquad (3.2.14)$$

where $J(u)$, $0\le u\le 1$ is a squared-integrable weight function called a score function. This class of statistics is called *linear rank statistics*. It is an easy matter to show (using change-of-variable techniques) the following equivalent representations for $T_N(J)$.

$$\begin{aligned} T_N(J) &= (1/m)\sum_{j=1}^{m} J(R(j)/(N+1)) \\ &= \int_0^1 J[\tfrac{N}{(N+1)}\hat{D}_1(u)]du = \int_0^1 J[\tfrac{N}{(N+1)}u]d\hat{D}(u) \quad . \end{aligned} \qquad (3.2.15)$$

(See Parzen (1983) p. 11 for a proof). Note that the last representation on the right, i.e,

$$T_N(J) = \int_0^1 J[\tfrac{N}{(N-1)}u]d\hat{D}(u)$$

is of particular importance to the development of the asymptotic

distribution of $T_N(J)$ since it exhibits $T_N(J)$ as a linear functional of the process $\tilde{D}(u)$. One expects that the asymptotic distribution of $T_N(J)$ is that of the same linear functional of the process which is the limit of $\tilde{D}(u)$ (appropriately normalized).

Before turning to the development of the asymptotic distribution of $\tilde{D}(u)$, we define four more functions that make the presentation of the results more symmetric and easy to follow. Those functions are

$$\begin{aligned} D_F(u) &= FQ_H(u) \\ d_F(u) &= D_F'(u) \\ D_G(u) &= GQ_H(u) \\ d_G(u) &= D_G'(u) \quad . \end{aligned}$$

Note that $D_F(u)=D(u)$ and $D_G(u)=(1/(1-\lambda))(u-\lambda D(u))$. The introduction of the new functions is indeed not necessary and is used only for convenience.

3.3 Asymptotic Distribution of $\tilde{D}(u)$

In this section we develop the asymptotic distribution of $\tilde{D}(u)$ for the case described in Section 3.1, i.e, when the two samples come from two independent time series satisfying the conditions for the weak convergence of the coresponding empirical processes. Let $\tilde{\alpha}_F$ and $\tilde{\alpha}_G$ be the empirical processes corresponding to $\{X\}$ and $\{Y\}$ respectively, i.e,

$$\tilde{\alpha}_F(u) = m^{1/2}[\tilde{F}Q_F(u)-u] \quad , \; 0\le u\le 1 \tag{3.3.1}$$

and

$$\tilde{\alpha}_G(u) = n^{1/2}[\tilde{G}Q_G(u)-u] \quad , \; 0\le u\le 1 \; . \tag{3.3.2}$$

Then we assume that

$$\tilde{\alpha}_F \xrightarrow{d} \alpha_F \tag{3.3.3}$$

and

$$\tilde{\alpha}_G \xrightarrow{d} \alpha_G \; , \tag{3.3.4}$$

where α_F and α_G are zero mean Gaussian processes with covariance kernels given, respectively, by

$$K_{\alpha_F}(u,v) = \mathrm{Cov}\{\alpha_F(u),\alpha_F(v)\} = K_0(u,v) + K_F(u,v) \; , \tag{3.3.5}$$

$$K_{\alpha_G}(u,v) = \mathrm{Cov}\{\alpha_G(u),\alpha_G(v)\} = K_0(u,v) + K_G(u,v) \; , \tag{3.3.6}$$

where

$$K_0(u,v) = u\wedge v-uv \quad , \tag{3.3.7}$$

$$K_F(u,v) = 2\sum_1^{\infty}[B_k(u,v;F)-uv] \; , \tag{3.3.8}$$

$$K_G(u,v) = 2\sum_1^{\infty}[B_k(u,v;G)-uv] \; , \tag{3.3.9}$$

and $B_k(\cdot,\cdot;F)$ and $B_k(\cdot,\cdot;G)$ are the dependence DFs of lag k for the processes $\{X\}$ and $\{Y\}$ respectively.

Next define the *comparison processes* $\tilde{\gamma}$ and $\tilde{\delta}$ by

$$\tilde{\gamma}(u) = N^{1/2}[\tilde{F}\tilde{Q}_H(u)-D(u)] \; , \quad 0\le u\le 1 \; , \tag{3.3.10}$$

and

$$\tilde{\delta}(u) = N^{1/2}[\tilde{D}(u)-D(u)] \; , \quad 0\le u\le 1 \; . \tag{3.3.11}$$

Now, using the explicit representations for $\tilde{F}\tilde{Q}_H$ and $\tilde{D}$ as given in 3.2.11 and 3.2.12, it is easy to see that

$$\sup\{|\tilde{\gamma}(u)-\tilde{\delta}(u)|:0\le u\le 1\} \le \lambda_N^{-1/2}\cdot m^{-1/2} \quad . \tag{3.3.12}$$

Hence, the two comparison processes have the same limiting behavior. Therefore, asymptotic results about $\tilde{\gamma}$ can be immediately applied to $\tilde{\delta}$.

Pyke and Shorack (1968) have studied the asymptotic behavior of $\tilde{\gamma}$ for the independence case and Fears and Mehra (1974) have extended their results to include the case of mixing processes. Here we sketch the main steps in the development of these results. Using Lemma 3.1 of Pyke and Shorack (1968) we have

$$\begin{aligned}\tilde{\gamma}(u) = (1-\lambda_N)\{\lambda_N^{-1/2}\tilde{d}_G(u)\tilde{\alpha}_F(F\tilde{Q}_H(u)) \\ -(1-\lambda_N)^{-1/2}\tilde{d}_F(u)\tilde{\alpha}_G(G\tilde{Q}_H(u))\} + \hat{R}(u) ,\end{aligned} \tag{3.3.13}$$

where

$$\tilde{d}_F(u) = [D_F(H\tilde{Q}_H(u))-D_F(u)]/[H\tilde{Q}_H(u)-u] , \tag{3.3.14}$$

$$\tilde{d}_G(u) = [D_G(H\tilde{Q}_H(u))-D_G(u)]/[H\tilde{Q}_H(u)-u] , \tag{3.3.15}$$

$$\tilde{R}(u) = \tilde{d}_F(u)N^{1/2}[\tilde{H}\tilde{Q}_H(u)-u] \quad . \tag{3.3.16}$$

Note that only algebraic manipulations are involved in the derivation of the above representation for $\tilde{\gamma}$, so that it is valid under any dependence structures of the underlying processes $\{X\}$ and $\{Y\}$. Next observe that $\tilde{d}_F$ and $\tilde{d}_G$ are related by the equation

$$\lambda_N\tilde{d}_F(u) + (1-\lambda_N)\tilde{d}_G(u) = 1 , \tag{3.3.17}$$

and since D_F and D_G are nondecreasing, $\tilde{d}_F$ and $\tilde{d}_G$ are nonnegative. Hence, we immediately have

$$|\tilde{d}_F(u)| \le \lambda_N^{-1} \quad 0\le u\le 1 \ , \tag{3.3.18}$$

and

$$|\tilde{d}_G(u)| \le (1-\lambda_N)^{-1} \quad 0\le u\le 1 \ . \tag{3.3.19}$$

Also, since

$$|\tilde{H}\tilde{Q}_H(u)-u| \le N^{-1} \quad 0\le u\le 1 \ , \tag{3.3.20}$$

we have

$$|\tilde{R}(u)| \le \lambda_N^{-1}N^{-1/2} \quad 0\le u\le 1 \ . \tag{3.3.21}$$

Next define

$$\begin{aligned}\gamma(u) = \delta(u) = (1-\lambda)\{&\lambda^{-1/2}d_G(u)a_F(D_F(u)) \\ &-(1-\lambda)^{-1/2}d_F(u)a_G(D_G(u))\} \ , \quad 0\le u\le 1 \ ,\end{aligned} \tag{3.3.22}$$

and observe that γ is the natural limit of $\tilde{\gamma}$ whenever convergence occurs. Using triangle inequality techniques (for the metrics involved) and the bounds in (3.3.18)-(3.3.21) above, the proof of the convergence of $\tilde{\gamma}$ to γ is essentially reduced to showing the appropriate convergences of $\tilde{d}_F(u)$ and $\tilde{d}_G(u)$ to $d_F(u)$ and $d_G(u)$ (resp.) and of $\tilde{a}_F(F\tilde{Q}_H(u))$ and $\tilde{a}_G(G\tilde{Q}_H(u))$ to $a_F(D_F(u))$ and $a_G(D_G(u))$ (resp.). Pyke and Shorack (1968) (for the independence case) and Fears and Mehra (1974) (for the dependence case) provide the detailed proofs of these convergences. Consequently, we have

$$\tilde{\gamma} \xrightarrow{d} \gamma \tag{3.3.23}$$

which implies

$$\tilde{\delta} \xrightarrow{d} \delta \ . \tag{3.3.24}$$

Observe that δ is a zero mean Gaussian process with covariance kernel

$$\begin{aligned} K_\delta(u,v) &= \mathrm{Cov}[\delta(u),\delta(v)] \\ &= (1-\lambda)^2\{\lambda^{-1}d_G(u)d_G(v)K_{a_F}(D_F(u),D_F(v)) \\ &\quad + (1-\lambda)^{-1}d_F(u)d_F(v)K_{a_G}(D_G(u),D_G(v))\} \ . \end{aligned} \tag{3.3.25}$$

3.4 The Asymptotic Distribution of $T_N(J)$

R-call that

$$T_N(J) = \int_0^1 J(\tfrac{N}{(N+1)}u)d\tilde{D}(u) \ , \tag{3.4.1}$$

and observe that the assumed continuity of J implies the asymptotic equivalence of $T_N(J)$ and

$$T_N^*(J) = \int_0^1 J(u)d\tilde{D}(u) \ . \tag{3.4.2}$$

Next define

$$\begin{aligned} \Delta_N(J) &= N^{1/2}\{T_N^*(J)-\int_0^1 J(u)dD(u)\} \\ &= \int_C^1 J(u)d\tilde{\delta}(u) \end{aligned} \tag{3.4.3}$$

and observe that

$$\tilde{\delta} \xrightarrow{d} \delta \tag{3.4.4}$$

implies

$$\Delta_N(J) \xrightarrow{d} \Delta(J) = \int_0^1 J(u)d\delta(u) \quad . \tag{3.4.5}$$

Further, $\Delta(J)$ is $N(0,\sigma_J^2)$, where

$$\sigma_J^2 = \int_0^1\int_0^1 J(u)J(v)dK_\delta(u,v) \quad . \tag{3.4.6}$$

One can also consider the joint convergence of $(\Delta_N(J_1),\ldots,\Delta_N(J_p))$ to $(\Delta(J_1),\ldots,\Delta(J_p))$, where the latter is a zero mean p-variate normal vector with covariance matrix

$$\begin{aligned} \sigma_{rs} &= \mathrm{Cov}(\Delta(J_r),\Delta(J_s))) \\ &= \int_0^1\int_0^1 J_r(u)J_s(v)dK_\delta(u,v) \ . \end{aligned} \tag{3.4.7}$$

A careful evaluation of the last integral leads to the expression

$$\sigma_{rs} = C_1(J_r,J_s) + C_3(J_r,J_s) - C_2(J_r,J_s) + C_4(J_r,J_s) \ , \tag{3.4.8}$$

where the terms on the right hand are defined by (3.4.9)-(3.4.19):

$$\begin{aligned} C_1(J_r,J_s) = (1-\lambda)^2\int_0^1 & J_r(u)J_s(u)[\lambda^{-1}d_G^2(u)d_F(u) \\ &+ (1-\lambda)^{-1}d_F{}^2(u)d_G(u)]du \quad , \end{aligned} \tag{3.4.9}$$

$$C_3(J_r,J_s) = (1-\lambda)^2\int_0^1\int_0^1 J_r(u)J_s(v)$$

$$\{\lambda^{-1}[A_1(u,v;G,G,F) + A_2(u,v;G,G,F)]$$

$$+(1-\lambda)^{-1}[A_1(u,v;F,F,G) + A_2(u,v;F,F,G)]\}dudv \quad (3.4.10)$$

where

$$A_1(u,v;G,G,F) = d_G'(u)d_G'(v)D_F(u\wedge v) \;, \quad (3.4.11)$$

$$A_2(u,v;G,G,F) = d_G'(u\vee v)d_G(u\wedge v)d_F(u\wedge v) \;, \quad (3.4.12)$$

$$C_2(J_r,J_s) = (1-\lambda)^2[\lambda^{-1}A_3(J_r,G,F)A_3(J_s,G,F)$$

$$+ (1-\lambda)^{-1}A_3(J_r,F,G)A_3(J_s,F,G)] \quad (3.4.13)$$

where

$$A_3(J_r,G,F) = \int_0^1 J_r(u)\{d_G(u)D_F(u)\}'du \;, \quad (3.4.14)$$

$$C_4(J_r,J_s) = (1-\lambda)^2\int_0^1\int_0^1 J_r(u)J_s(v)$$

$$\{\lambda^{-1}[A_4(u,v;G,F) + A_5(u,v;G,F)$$

$$+ A_6(u,v;G,F) + A_7(u,v;G,F)]$$

$$+ (1-\lambda)^{-1}[A_4(u,v;F,G) + A_5(u,v;F,G)$$

$$+ A_6(u,v;F,G) + A_7(u,v;F,G)]\}dudv \quad (3.4.15)$$

where

$$A_4(u,v;G,F) = d_G'(u)d_G'(v)K_F(D_F(u),D_F(v)) \ , \qquad (3.4.16)$$

$$A_5(u,v;G,F) = d_G'(u)d_G(v)d_F(u)K_F^{(2)}(D_F(u),D_F(v)) \ , \qquad (3.4.17)$$

$$A_6(u,v;G,F) = d_G(u)d_G'(v)d_F(u)K_F^{(1)}(D_F(u),D_F(v)) \ , \qquad (3.4.18)$$

$$A_7(u,v;G,F) = d_G(u)d_G(v)d_F(u)d_F(v)$$

$$\cdot 2\sum_1^\infty [b_k(D_F(u),D_F(v);F)-1] \qquad (3.4.19)$$

and $K^{(1)}$ and $K^{(2)}$ are the partial derivatives of K w.r.t the 1^{st} and 2^{nd} arguments, respectively. Recall that b_k is the dependence density function corresponding to the dependence DF B_k. Clearly, expression (3.4.8) is not very easy to work with (although it is completely symmetric w.r.t F and G).

Under the null hypothesis, H_0:F≡G, we have

$$D_F(u) = D_G(u) = u \ ,$$

$$d_F(u) = d_G(u) = 1 \ , \qquad (3.4.20)$$

$$d_F'(u) = d_G'(u) = 0 \ .$$

Consequently, the expression for σ_{rs} reduces to

$$\sigma_{rs} = \frac{1-\lambda}{\lambda}V(J_r,J_s)$$

$$+ 2(1-\lambda)^2\sum_1^\infty[\lambda^{-1}V_k(J_r,J_s,F) + (1-\lambda)^{-1}V_k(J_r,J_s,G)] \qquad (3.4.21)$$

where

$$V(J_r,J_s) = \int_0^1 J_r(u)J_s(u)du - \int_0^1 J_r(u)du\int_0^1 J_s(u)du$$

$$= \mathrm{Cov}[J_r(F(X(i))),J_s(F(X(i)))]$$

$$= \mathrm{Cov}[J_r(G(Y(j))),J_s(G(Y(j)))] \qquad (3.4.22)$$

$$V_k(J_r,J_s;F) = \int_0^1\int_0^1 J_r(u)J_s(v)[b_k(u,v;F)-1]dudv$$

$$= \mathrm{Cov}[J_r(F(X(i))),J_s(F(X(i+k)))] \quad . \qquad (3.4.23)$$

Consequently, rearranging terms, we have

$$\sigma_{rs} = \frac{1-\lambda}{\lambda}\{(1-\lambda)[V(J_r,J_s) + 2\sum_1^\infty V_k(J_r,J_s;F)]$$

$$+ \lambda[V(J_r,J_s) + 2\sum_1^\infty V_k(J_r,J_s;G)]\} \quad . \qquad (3.4.24)$$

For the case r=s, this can be reduced even more. We then have

$$\sigma_J^2 = \frac{1-\lambda}{\lambda}\mathrm{Var}(J(U))\{(1-\lambda)[1 + 2\sum_1^\infty\rho(k;J,F)]$$

$$+ \lambda[1 + 2\sum_1^\infty\rho(k;J,G)]\}$$

$$= \frac{1-\lambda}{\lambda}\mathrm{Var}(J(U))[(1-\lambda)f(0;J,F) + \lambda f(0;J,G)] \ , \qquad (3.4.25)$$

where $\mathrm{Var}(J(U)) = \int_0^1 J^2(u)du - [\int_0^1 J(u)du]^2$,

$\rho(\cdot;J,F)$ is the correlation function of the time series $\{J(F(X(i))):i\epsilon Z\}$,

$f(\cdot;J,F)$ is the spectral density function that corresponds to $\rho(\cdot;J,F)$, i.e

$$f(\omega;J,F) = 1+2\sum_{1}^{\infty}\rho(k;J,F)\cos(2\pi\omega k) \ , \ 0\le\omega\le 1 \ , \quad (3.4.26)$$

and $\rho(\cdot;J,G)$, $f(\cdot;J,G)$ are the corresponding quantities for the time series $\{J(G(Y(j))):j\epsilon Z\}$.

Remarks :

1) The terms C_1, C_2 and C_3 in the general expression for σ_{rs} are xactly the same under independence, while the last term, C_4, is contributed by the dependence structure of the two processes $\{X\}$ and $\{Y\}$. This term vanishes under independence since then $K_F = K_G = 0$.

2) Under H_0, C_3 vanishes, C_1 reduces to

$$\frac{1-\lambda}{\lambda}\int_0^1 J_r(u)J_s(u)du \quad (3.4.27)$$

and C_2 reduces to

$$\frac{1-\lambda}{\lambda}\int_0^1 J_r(u)du\int_0^1 J_s(u)du \quad (3.4.28)$$

so that $V(J_r,J_s)$ is the covariance under independence while the last term, $V_k(J_r,J_s,\cdot)$, is contributed by the dependence structures.

3) Note that $b_k(\cdot,\cdot;F)$ is the joint density of the RVs $F(X(i))$ and $F(X(i+k))$ (any $i\epsilon Z$), and 1 is the product of the marginal densities of these two RVs and hence

$$\int_0^1\int_0^1 J_r(u)J_s(v)[b_k(u,v;F)-1]dudv$$

$$= Cov[J_r(F(X(i))),J_s(F(X(i+k)))] \ . \quad (3.4.29)$$

3.5 Interpretations

The formula for the asymptotic variance of $T_N(J)$, obtained in the last section, is of particular interest since it emphasizes and isolates the effect of dependence and, moreover, it expresses that effect in an interpretable and estimable form. The first thing to observe is that the term

$$\sigma_0^2(J) = \frac{1-\lambda}{\lambda}\mathrm{Var}[J(U)] \tag{3.5.1}$$

is the variance under independence, so that the multiplicative term on the right, i.e

$$\sigma_D^2(J) = [(1-\lambda)f(0;J,F)+\lambda f(0;J,G)] \tag{3.5.2}$$

is the effect of the dependence structures of $\{X\}$ and $\{Y\}$. Next observe that this term is a weighted average of the effects of each of the time series $\{X\}$ and $\{Y\}$, while these effects are expressed as the values at zero of the spectral density functions of the time series $\{J(F(X))\}$ and $\{J(G(Y))\}$. The fact that the dependence effect of each series is summarized in a single value (as opposed to being expressed as an infinite sum) makes it easier to analyze that effect both qualitatively and quantitatively. Observe that whether the dependence effect is greater than or smaller than 1 determines whether it increases or decreases the variance of the statistic relative to independence. Now, the value of a spectral density function at 0 has a special meaning. Since

$$f(0) = 1 + 2\sum_1^\infty \rho(k) \ , \tag{3.5.3}$$

where f is the spectral density function of a time series with correlation function $\rho(\cdot)$, $f(0)>1$ occurs when $\sum_{1}^{\infty}\rho(k) > 0$ and this last condition can be viewed as a kind of positive dependence in the time series. Similarly, $f(0)<1$ can be viewed as a kind of negative dependence. When $T_N(J)$ is used to test the hypothesis $H_0{:}F\equiv G$ versus the general alternative $H_1{:}F\neq G$, a critical region of the form $|\Delta_N(J)|>c_N(\alpha)$ (with $D(u)=u$ in $\Delta_N(J)$) is usually used with $c_N(\alpha)\longrightarrow\sigma_0(J)Q_\Phi(1-2\alpha)$. If dependence is present and is not taken into account, it can affect the size of the test. If $\sigma_D^2(J)>1$ then the test will reject too often (i.e actual α > nominal α) while the opposite will occur when $\sigma_D^2<1$.

Remark: The issue of positive dependence has been discussed in the literature. One of the common definitions of positive dependence is the following. The RVs X and Y are said to be positively dependent if

$$E[h(X)h(Y)] \geq 0 \text{ for all } h \text{ with } E|h(X)h(Y)| < \infty. \tag{3.5.4}$$

Gleser and Moore (1983) applied this condition to any pair $(X(i),X(j))$ of a stationary time series and showed that tests of fit may reject the null hypothesis of fit too often under dependence when independence is assumed. This is shown to hold for wide range of tests including chi-squared type tests and tests based on the EDF. We want to mention here that their condition is much stronger than the condition $f(0;X)>1$. First, the condition $E[h(X(1))h(X(1+k))]\geq 0$ for any function h is stronger than $\rho_X(k)\geq 0$ (although the two conditions coincide for Gaussian processes) and secondly, $\rho_X(k)\geq 0$ for

all k is much stronger than $\sum_1^{\infty}\rho_X(k)\geq 0$. Moreover, their conclusion, being completely qualitative, is weaker than the one given here which expresses the effect of the dependence explicitly and even suggests a way to estimate it. We should mention, however, that the result here is applicable only for one specific class of statistics, namely, the linear rank statistics for the two-sample problem, while their results and methods are applicable for a much wider range of statistics.

3.6 Estimation

In the previous section we have discussed the effect of dependence on various test statistics when the tests are formed assuming independence. Here we are going to discuss a solution to the problem, i.e, what can be done when one is willing to take into account the dependence structure. The first thing to observe is that the asymptotic normality of the test statistic continues to hold (under the appropriate conditions of smoothness and mixing), so that the only other thing needed in order to determine a critical region is an estimable expression for the asymptotic variance of the test statistic. The formula for the variance discussed here is almost directly estimable. Since procedures for spectral density estimation are easily available, the only problem is that the time series $\{J(F(X(j))):j\epsilon Z\}$ and $\{J(G(Y(i))):i\epsilon Z\}$ are not observable since the functions F and G are unknown (note that the function J is completely specified for any specific statistic at hand). However, we have a natural estimator for a DF, namely, the sample DF, so we

can take the observed time series $\{\tilde{F}(X(j)):j=1,\ldots,m\}$ and $\{\tilde{G}(Y(i)):i=1,\ldots,n\}$ as estimators of finite realizations from the time series $\{F(X(j)):j\epsilon Z\}$ and $\{G(Y(i)):i\epsilon Z\}$ respectively. Note that by the Glivenko-Cantelli Lemma,

$$\max\{|\tilde{F}(X(j))-F(X(j))|:j=1,\ldots,m\} \le \sup\{|\tilde{F}(x)-F(x)|:x\epsilon \mathbb{R}\} \xrightarrow{as} 0 \quad (3.6.1)$$

(with similar result for Y) so that the above estimating samples do make sense. (Note that although the Glivenko-Cantelli Lemma was originally proved for independent RVs it is easy to show that it also holds for ϕ-mixing, stationary time series with $\sum_1^\infty \phi(k)<\infty$).

Remark: If we denote by $R_X(j)$ the rank of $X(j)$ within the X-sample and, similarly, by $R_Y(i)$ the rank of $Y(i)$ within the Y-sample, then we have

$$\tilde{F}(X(j)) = R_X(j)/m \quad \text{and} \quad \tilde{G}(Y(i)) = R_Y(i)/n \ . \quad (3.6.2)$$

Consequently, we call the time series $\{\tilde{F}(X(j))\}$ and $\{\tilde{G}(Y(i))\}$ the *rank transform* of the time series $\{X(j)\}$ and $\{Y(i)\}$ respectively. These rank transform time series appear to be very interesting in their own right. In particular, the relationships between the original time series and its rank transform version are unknown yet. We will try to explore some of those relationships in the next chapter.

3.7 Applications

Suppose that the processes $\{X(i):i\epsilon Z\}$ and $\{Y(j):j\epsilon Z\}$ are Gaussian time series that satisfy the conditions for the convergence of $\tilde{\delta}$ to δ. Let ρ_X and ρ_Y be the correlation functions of $\{X\}$ and $\{Y\}$ respectively. Then the dependence densities $b_k(\cdot,\cdot;F)$ and $b_k(\cdot,\cdot;G)$ have the same general form given by

$$b_k(u,v;\Phi) = (1-\rho_X^2(k))^{-1/2}\exp\{-.5(1-\rho_X^2(k))^{-1} \cdot[\rho_X^2(k)Q_\Phi^2(u)+\rho_X^2(k)Q_\Phi^2(v)-2\rho_X(k)Q_\Phi(u)Q_\Phi(v)]\} \tag{3.7.1}$$

where $Q_\Phi=\Phi^{-1}$ is the quantile function of a N(0,1) RV.

We will now try to evaluate the asymptotic variance of some common linear rank statistics.

a) The Wilcoxon Statistic, W .

For the Wilcoxon statistic we have

$$J_W(u) = u . \tag{3.7.2}$$

Hence,

$$\mathrm{Var}[J_W(U)] = 1/12 . \tag{3.7.3}$$

Next we need to evaluate $\rho(k;J_W,\Phi)$. We have

$$\rho(k;J_W,\Phi) = 12\mathrm{Cov}[\Phi(Z_1),\Phi(Z_2)] = 12E[\Phi(Z_1)\Phi(Z_2)] - 3 , \tag{3.7.4}$$

where Z_1 and Z_2 are N(0,1) RVs with correlation coefficient $\rho_X(k)$ (for $\{X\}$) or $\rho_Y(k)$ (for $\{Y\}$). In what follows we are using results

given in Owen (1980) to evaluate integrals for the normal distribution. For simplicity, we will denote the correlation between Z_1 and Z_2 simply by ρ. We have

$$E[\Phi(Z_1)\Phi(Z_2)] = E\{\Phi(Z_1)E[\Phi(Z_2)|Z_1]\} \ . \tag{3.7.5}$$

Now, $Z_2|Z_1=z$ is a $N(\rho z, 1-\rho^2)$ RV. Hence (using eq. 10,100.8 p. 403 Owen (1980)),

$$\begin{aligned} E[\Phi(Z_2)|Z_1=z] &= \int_{\mathbb{R}} \Phi(t)d\Phi[(t-\rho z)/(1-\rho^2)^{1/2}] \\ &= \int_{\mathbb{R}} \Phi[\rho z+(1-\rho^2)^{1/2}t]d\Phi(t) \\ &= \Phi[\rho z/(2-\rho^2)^{1/2}] \quad . \end{aligned} \tag{3.7.6}$$

So (using eqs. 2,010.6 and 2,010.7 p. 400 Owen(1980)),

$$\begin{aligned} E[\Phi(Z_1)\Phi(Z_2)] &= \int_{\mathbb{R}} \Phi(t)\Phi[\rho t/(1-\rho^2)^{1/2}]d\Phi(t) \\ &= 1/4 + (1/2\pi)\tan^{-1}[\rho/(4-\rho^2)^{1/2}] \\ &= 1/4 + (1/2\pi)\sin^{-1}(\rho/2) \quad . \end{aligned} \tag{3.7.7}$$

Hence

$$\rho(k;J_W,\Phi) = (6/\pi)\sin^{-1}(\rho/2) \ , \tag{3.7.8}$$

where, again, ρ represents either $\rho_X(k)$ or $\rho_Y(k)$. Consequently, we have

$$\begin{aligned} \mathrm{Var}[\Delta(J_W)] = \tfrac{1-\lambda}{\lambda}(1/12)\{&(1-\lambda)[1+(12/\pi)\sum_1^{\infty}\sin^{-1}(\rho_X(k)/2)] \\ &+ \lambda[1+(12/\pi)\sum_1^{\infty}\sin^{-1}(\rho_Y(k)/2)]\} \ . \end{aligned} \tag{3.7.9}$$

b) The Normal Score Test, N .

In this case $J_N(u) = Q_\Phi(u)$. Hence

$$J_N(F(X(i))) = X(i) \quad \text{and} \quad J_N(G(Y(j))) = Y(j) \ . \tag{3.7.10}$$

So, we immediately have

$$\mathrm{Var}[\Delta(J_N)] = \tfrac{1-\lambda}{\lambda}[(1-\lambda)f(0;X) + \lambda f(0;Y)] \ . \tag{3.7.11}$$

c) The Median Test, M .

In this case $J_M(u) = \mathrm{sgn}(u-.5)$. Hence

$$\mathrm{Var}[J_M(U)] = 1 \ , \tag{3.7.12}$$

and

$$\begin{aligned}
\rho(k;J_M,F) &= \mathrm{Corr}[J_M(F(X(i))),J_M(F(X(i+k)))] \\
&= E\{\mathrm{sgn}[F(X(i))-.5]\mathrm{sgn}[F(X(i+k))-.5]\} \\
&= E[\mathrm{sgn}(X(i)-\mathrm{med}(X))\mathrm{sgn}(X(i+k)-\mathrm{med}(X))] \\
&= \Pr[(X(i)-\mathrm{med}(X))(X(i+k)-\mathrm{med}(X)) > 0] \\
&\quad - \Pr[(X(i)-\mathrm{med}(X))(X(i+k)-\mathrm{med}(X)) < 0] \\
&= 2\Pr[(X(i)-\mathrm{med}(X))(X(i+k)-\mathrm{med}(X)) > 0] - 1 \\
&= 2\{\Pr[X(i) < \mathrm{med}(X),X(i+k) < \mathrm{med}(X)] \\
&\quad + \Pr[X(i) > \mathrm{med}(X),X(i+k) > \mathrm{med}(X)]\} - 1
\end{aligned} \tag{3.7.13}$$

where $\mathrm{med}(X)=Q_F(.5)$ is the median of X. Note that this is a familiar

measure of association, it is usually denoted by q and is discussed in Blomqvist (1950). For jointly normal RVs, med(X)=E(X(i)) and, also, the two probabilities on the last expression above are equal. Hence,

$$\rho(k;J_M,F) = 4\Pr(Z_1<0,Z_2<0) - 1 , \tag{3.7.14}$$

where (Z_1,Z_2) is a standard bivariate normal RV with correlation coefficient $\rho_X(k)$. Hence (using eq. 3.5 p. 416 and eq. 2.2 p. 414 Owen (1980)),

$$\begin{aligned}\rho(k;J_M,F) &= 4\{1/2 - (1/\pi)\tan^{-1}[((1-\rho)/(1+\rho))^{1/2}]\} - 1 \\ &= 1 - (4/\pi)\tan^{-1}[((1-\rho)/(1+\rho))^{1/2}] \\ &= (2/\pi)\sin^{-1}(\rho) . \end{aligned} \tag{3.7.15}$$

Consequently,

$$\begin{aligned}\mathrm{Var}[\Delta(J_M)] = \tfrac{1-\lambda}{\lambda}\{&(1-\lambda)[1 + (4/\pi)\sum_1^{\infty}\sin^{-1}(\rho_X(k))] \\ &+ \lambda[1 + (4/\pi)\sum_1^{\infty}\sin^{-1}(\rho_Y(k))]\} . \end{aligned} \tag{3.7.16}$$

3.8 Empirical Examples

In this section we are going to analyze briefly two data sets, applying some of the procedures discussed above. The first data set consists of two samples of size 32 each. It represents the reflectance of energy from a mineral in 32 disjoint bands of frequency. Each sample represents a different mineral and the question is whether these two minerals differ in the way they reflect

energy. Clearly, the two samples are independent, but the observations within each sample are probably not independent.

Remark: There is always the general problem of applying asymptotic results to finite samples. Practically, almost any sample can be exposed to the computer programs that carry out the computations of a statistical procedure. The question is how to interpret the results. Our approach is to use the statistical procedures to learn about the data and, at the same time, to use the data to learn about the behavior of the statistical procedure. Consequently, when we have a small sample, we use the statistical procedure as an exploratory tool and do not draw definite conclusions.

The analysis of the second data set demonstrates an application of the two-sample procedure in a time series situation which can be regarded as a check for stationarity of the time series. To test whether the marginal distributions of the components of a time series remain unchanged as time changes, one approach is to take the two end parts of the time series and expose them to the two-sample procedure. Since the observations within each sample are not independent, the effect of dependence needs to be taken into account. Note that the statistical procedure described here is not strictly adequate for this application since the assumed independence of the twc samples is not completely satisfied. However, when the two samples come from far apart portions of the time series, we can still use our exploratory approach to gain insight into the problem. The remark above about applying asymptotic results to finite samples applies also here.

The second data set we analyze here was formed from the well-known International Airlines time series. Since this time series has a clearly recognized trend and twelve month cycle in it we follow the commonly used practice and transform the data by taking the 12^{th} difference of the log of the original data. The resulting time series has 132 observations, so we formed the two samples by taking the first and last thirds. This gives two samples of size 44 each which we analyze as our second example below.

a) Computational Formulas

Given two samples $X(1),\ldots,X(m)$ and $Y(1),\ldots,Y(n)$, we first order each of them and denote by $X(1:m),\ldots,X(m:m)$ and $Y(1:n),\ldots,Y(n:n)$ the resulting order statistics vectors. Then the ranks $R_1,\ldots,R_m$ are calculated by

$$R_j = j + \sum_{i=1}^{n} I[Y(i:n) \le X(j:m)] \quad , \; j=1,\ldots,m \quad . \tag{3.8.1}$$

Next, the comparison DF $\tilde{D}(u)$ is represented by an m+2 ordinate vector $\tilde{D}_j$ and by an m+2 abscissa vector $\tilde{U}_j$ defined by

$$\begin{aligned} \tilde{D}_j &= (j-1)/m \quad , \; j=1,\ldots,m+1 \\ \tilde{D}_{m+2} &= 1 \\ \tilde{U}_{j+1} &= R_j/N \quad , \; j=0,\ldots,m+1 \end{aligned} \tag{3.8.2}$$

where $N=m+n$, $R_0=0$ and $R_{m+1}=N$. Then we plot $\tilde{D}_j$ versus $\tilde{U}_j$ joining the points by straight lines. Next we calculate three linear rank

statistics of the form

$$T_N(J) = (1/m) \sum_{j=1}^{m} J(R_j/(N+1)) \quad , \tag{3.8.3}$$

where

$J(u) = J_W(u) = u$ for the Wilcoxon statistic,

$J(u) = J_N(u) = Q_\Phi(u)$ for the normal score statistic,

$J(u) = J_M(u) = \mathrm{sgn}(u-.5)$ for the median statistic.

Then we calculate the mean and variance under independence, and under the null hypothesis, for each of the statistics using the following formulas.

$$E[T_N(J)] = \frac{1}{N} \sum_{k=1}^{N} J(\tfrac{k}{N+1}) \quad , \tag{3.8.4}$$

$$V[T_N(J)] = \frac{1}{N} \cdot \frac{n}{m} \cdot \frac{1}{N-1} \sum_{k=1}^{N} \{J(\tfrac{k}{N+1}) - E[T_N(J)]\}^2 \quad . \tag{3.8.5}$$

Note that

$$E[T_N(J)] \longrightarrow \mu(J) = \int_0^1 J(u)du \quad \text{and}$$

$$NV[T_N(J)] \longrightarrow \sigma_0^2 = \frac{1-\lambda}{\lambda} \int_0^1 [J(u)-\mu(J)]^2 du \quad \text{as } N \longrightarrow \infty \quad .$$

Next we normalize each statistic by

$$T_N^*(J) = \{T_N(J) - E[T_N(J)]\} / \{V[T_N(J)]\}^{1/2} \quad . \tag{3.8.6}$$

Note that under independence (and under the null hypothesis) $T_N^*(J) \xrightarrow{d} N(0,1)$. Next we estimate the effect of the dependence structure within each sample on the asymptotic variance of $T_N^*(J)$. This is done by estimating the spectral density of the time series

$\{J(\tilde{F}(X(j)))\}$ and $\{J(\tilde{G}(Y(i)))\}$ for each of the score functions J_W, J_N and J_M. The formulas for the spectral estimation procedures are described in the next chapter. After having estimates of $f(0;J,F)$ and $f(0;J,G)$, we calculate

$$\sigma_D^2(J) = (1-\lambda)f(0;J,F)+\lambda f(0;J,G) \tag{3.8.7}$$

where $\lambda=m/N$. Finally, we let $T_N{}^D(J)=T_N{}^*(J)/\sigma_D(J)$, and check whether $|T_N{}^D(J)|>Q_\Phi(1-\alpha/2)$ in order to approximate a level α test for the null hypothesis.

b) The Results

The plot of $\tilde{D}_j$ for the reflectance data is presented in Figure 1. A diagonal line representing the function $D(u)=u$ has been added to the plot to help comparing $\tilde{D}(u)$ to its theoretical value under the null hypothesis. This plot suggests that the two populations are different, the first one being stochastically smaller. We infer this from the fact that $\tilde{D}(u)>D(u)$ $0\leq u\leq 1$. Table 1 presents the numerical results for the three linear rank statistics: the Wilcoxon, the normal score and the median. The symbols at the top of each column in the table are those defined in the previous subsection. The values for $T_N{}^*(J)$ show that under independence the null hypothesis can be rejected with p-value less than .001 . However, when dependence is taken into account, the values for $T_N{}^D(J)$ show that the null hypothesis can no longer be rejected so safely and it might be that the autocorrelations within each sample caused the two samples to look so different.

The results for the second data set, the International Airlines data, lead to similar conclusions. The plot of $\tilde{D}_j$, presented in figure 2, suggests that the two samples are different, the first one being stochastically larger. Table 2 presents the numerical results for the three linear rank statistics. The values of $T_N^*(J)$ indicate that under independence the null hypothesis can be rejected with p-value less than .001 . But, again, the autocorrelations within each sample increase the p-value to around .05, so we have to be more careful in rejecting the null hypothesis.

Table 1. REFLECTANCE DATA. Linear Rank Statistics - Numerical Results.

Statistic	$T_N(J)$	$T_N^*(J)$	$f(0;J,F)$	$f(0;J,G)$	$\sigma_D^2(J)$	$T_N^D(J)$
Wilcoxon	.319	-5.049	6.295	5.427	5.861	-2.086
Normal	-.595	-4.997	6.011	5.278	5.644	-2.103
Median	-.500	-3.969	6.032	4.784	5.408	-1.707

Table 2. INTERNATIONAL AIRLINES DATA. Linear Rank Statistics - Numerical Results.

Statistic	$T_N(J)$	$T_N^*(J)$	$f(0;J,F)$	$f(0;J,G)$	$\sigma_D^2(J)$	$T_N^D(J)$
Wilcoxon	.635	4.415	2.950	4.819	3.885	2.240
Normal	.460	4.484	3.023	4.864	3.944	2.258
Median	.364	3.392	2.811	3.413	3.112	1.923

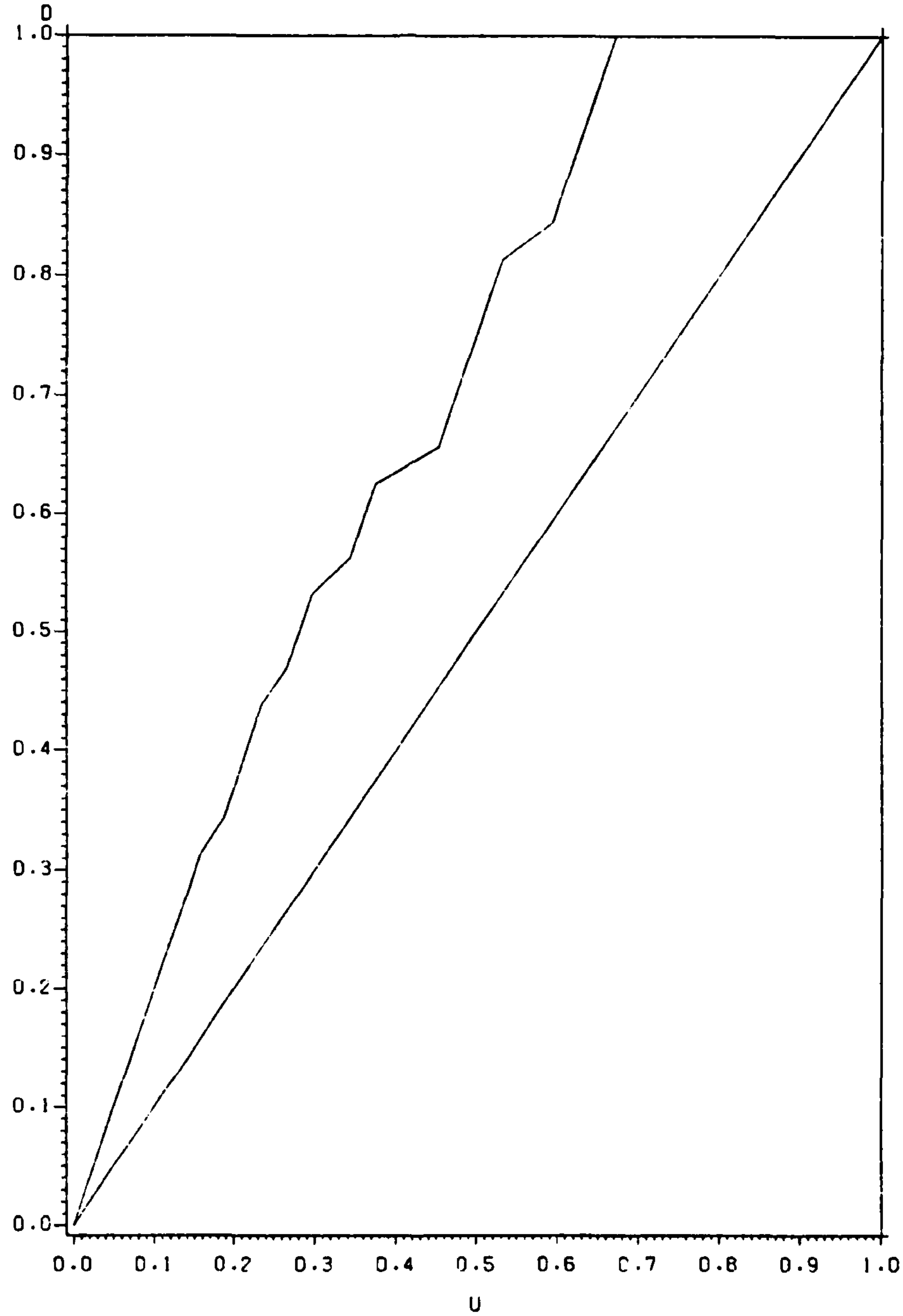


Figure 1. REFLECTANCE DATA

- The comparison DF $\tilde{D}(u)$ is the wiggly curve; it is compared with the uniform DF $D_0(u)=u$ given by the straight line.

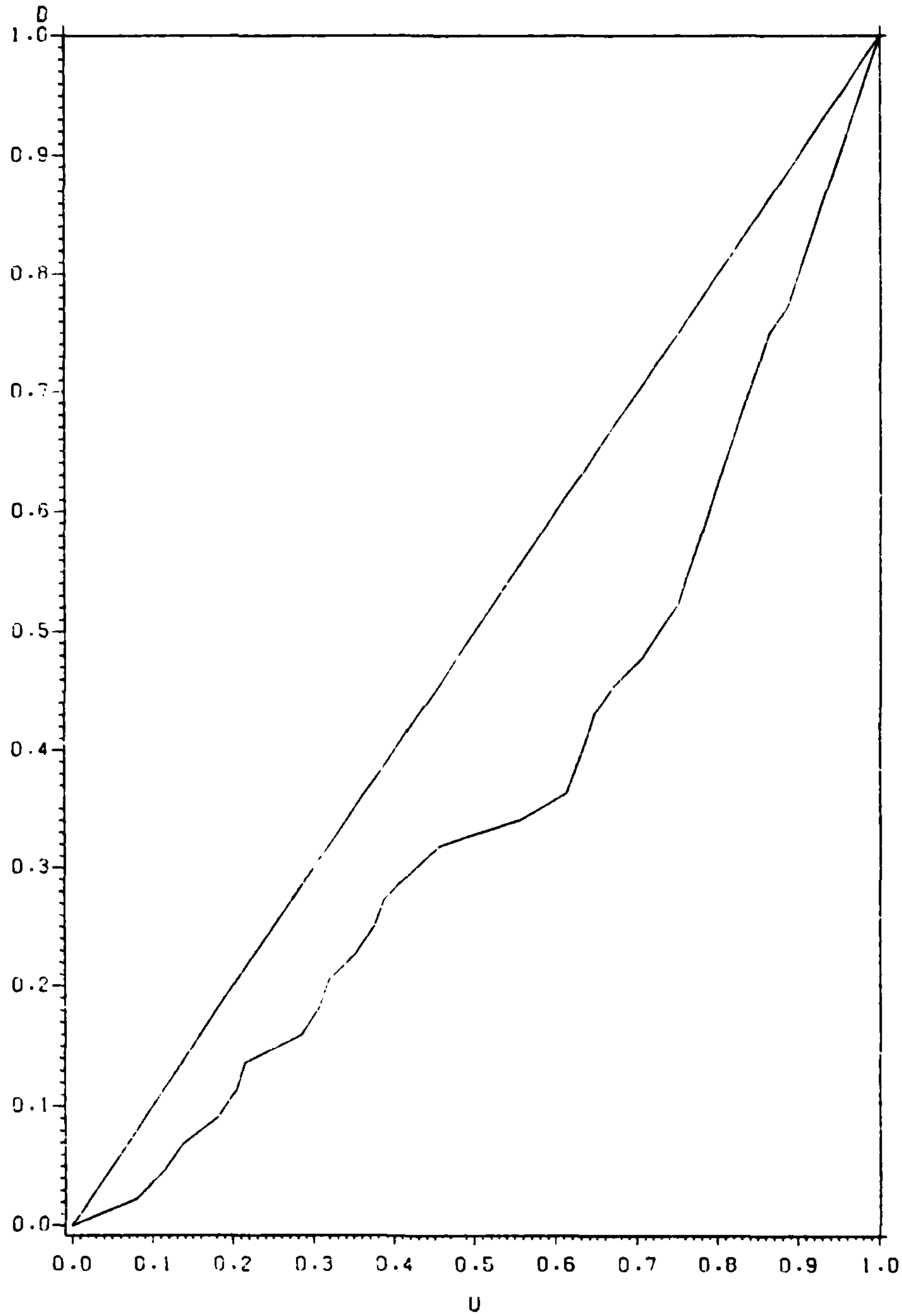


Figure 2. INTERNATIONAL AIRLINES DATA
- The comparison DF $\tilde{D}(u)$ is the wiggly curve; it is compared with the uniform DF $D_0(u)=u$ given by the straight line.

CHAPTER IV

RANK TRANSFORM SPECTRUM : EMPIRICAL TECHNIQUES

4.1 Introduction

For a stationary time series $\{Y(j):j\epsilon Z\}$ with a marginal DF F, we define the *probability integral transform* time series $\{U_Y(j):j\epsilon Z\}$ by

$$U_Y(j) = F(Y(j)) \ , \quad j\epsilon Z \ . \tag{4.1.1}$$

For a finite realization $Y(1),\ldots,Y(N)$ from $\{Y\}$, we define the *rank transform* time series $\tilde{U}_Y(j)$, $j=1,\ldots,N$ by

$$\tilde{U}_Y(j) = \tilde{F}(Y(j)) \ , \quad j=1,\ldots,N \ , \tag{4.1.2}$$

where $\tilde{F}$ is the EDF of the sample $Y(j)$, $j=1,\ldots,N$. As mentioned in the previous chapter, one can regard the sample $\tilde{U}_Y(j)$, $j=1,\ldots,N$ as an estimator of a sample from the time series $\{U_Y(j):j\epsilon Z\}$. Another way to look at the rank transform time series is as a general monotone transformation of $\{Y\}$.

In time series analysis, the theory and techniques used do not require assumptions on the marginal distribution of the RVs building the series (except of having finite second moments). However, the effect of the marginal distribution on the dependence structure is still of interest. One way to start gaining knowledge on this effect is to expose the rank transform time series to the time series analysis procedure used to analyze the original time series, and then compare the results. During this research work, the above procedure was repeated for several time series. Our general impression is that

the shape of the spectral density is invariant under the rank transformation and probably, under a general monotone transformation.

In this chapter, we are going to present these results for two time series: the Wolfer Sunspots data and the Critical Radio Frequencies data. In section 2 below we describe the computational formulas that were used to calculate the various spectral estimates. Section 3 contains the plots of these spectral estimates.

4.2 Computational Formulas

Let $Y(1),\ldots,Y(N)$ be the observed sample, and let $\tilde{Q}(u)$ be the sample quantile function calculated from that sample. Then, before entering the spectral analysis procedure, we standardize the sample by subtracting the median and dividing by twice the interquartile range, i.e, we take

$$Y^*(j) = [Y(j)-\tilde{Q}(.5)]/[2(\tilde{Q}(.75)-\tilde{Q}(.25))] \ , \quad j=1,\ldots,N \ . \qquad (4.2.1)$$

For convenience, we will keep denoting by $Y(j)$ the standardized version defined above. In addition to N, the sample size, we define two other vector sizes, denoted by NC and NF. NC is the number of lags for which correlations are calculated and is taken as $\min\{250,(N/2)\}$. NF is the number of equally spaced frequencies at which spectral quantities are calculated. This number is chosen as the smallest integer of the form $2^k 3^m 5^n$ which is greater than or equal N+NC. This choice enable us to calculate covariances by inverting the periodogram through the Discrete Fourier Transform (DFT), and also it makes the DFT routine work most efficiently.

First we calculate the unnormalized sample periodogram, f_N^*, by

$$f_N^*(\omega_p) = (1/N)\left|\sum_{j=1}^{N} Y(j)\exp[2\pi i(j-1)\omega_p]\right|^2 \quad , \tag{4.2.2}$$

where $\omega_p=(p-1)/NF$, $p=1,\ldots,NF$, $i=(-1)^{1/2}$, and $|z|^2$ is the squared modulus of a complex number z. The covariance function, $R(\cdot)$, is then calculated by taking the inverse DFT of the periodogram, i.e,

$$R(k) = (1/NF)\sum_{p=1}^{NF} f_N^*(\omega_p)\exp[-2\pi ik\omega_p] \quad , \quad k=0,1,\ldots,NC \quad . \tag{4.2.3}$$

Then we normalize the periodogram by

$$f_N(\omega_p) = f_N^*(\omega_p)/R(0) \quad , \tag{4.2.4}$$

and calculate the correlation function, $\rho(\cdot)$, by

$$\rho(k) = R(k)/R(0) \quad , \quad k=0,1,\ldots,NC \quad . \tag{4.2.5}$$

As can be seen above, we define spectral functions on the unit interval, so that they are even about ω=0.5. Also, to examine spectral functions we plot their logarithm, using a fixed scale axes with the abcsissa (representing the frequencies) running from 0 to 0.5 and the ordinate (representing the log spectra) running from -6.0 to +6.0 . The next step of the analysis is to plot the sample periodogram function, f_N. Next we calculate a spectral window smoothed periodogram using the Parzen window. The computational formulas for this estimator are as follows. The Parzen window function, w(k), is defined by

$$w(k) = \begin{cases} [1-6((k/NC)^2-(k/NC)^3)] & , k=0,1,\ldots,NC/2 \\ 2[1-(k/NC)]^3 & , k=NC/2,\ldots,NC \end{cases} \quad . \tag{4.2.6}$$

Then the smoothed periodogram, f_w, is calculated by

$$f_w(\omega_p) = 1+2\sum_{k=1}^{NC} w(k)\rho(k)\cos(2\pi k\omega_p) \quad . \tag{4.2.7}$$

This function is then plotted using the standard spectral plot described above. The last spectral estimator presented here is the autoregressive spectral density using the CAT order determination function. The computational formulas for this procedure are given below. First, the correlation function, $\rho(\cdot)$, is used to calculate the partial autocorrelation function, $pac(\cdot)$, by solving the Yule-Walker equations successively for orders 1,...,NC. For order k, the Yule-Walker equations are given by

$$V_k = \sum_{j=0}^{k} a_k(j)\rho(k)$$

and (4.2.8)

$$0 = \sum_{j=0}^{k} a_k(j)\rho(|j-m|) \quad , m=1,\ldots,k$$

where $a_k(0)=1$. These equations are solved for $a_k(1),\ldots,a_k(k)$ and for V_k using the given correlations $\rho(0),\rho(1),\ldots,\rho(k)$. The a_k's are the coefficients of the AR(k) model with these first k correlations and V_k is the residual variance for that model. After solving the equations for orders 1,...,NC, the partial autocorrelations are given by

$$pac(k) = a_k(k) \quad , k=1,\ldots,NC \quad . \tag{4.2.9}$$

The Yule-Walker equations are solved for successive orders using the

following recursive algorithm.

$$a_0(0) = 1$$

$$V_0 \quad = \rho(0) = 1$$

$$a_k(0) = 1$$

$$a_k(k) = \sum_{j=0}^{k-1} a_{k-1}(j)\rho(k-j)/V_{k-1}$$

$$a_k(j) = a_{k-1}(j)+a_k(k)a_{k-1}(k-j) \quad , \; j=1,\ldots,k-1$$

$$V_k \quad = \sum_{j=0}^{k} a_k(j)\rho(j) \quad .$$

The partial autocorrelations are then used to calculate the residual variances RV(·) using the formula

$$RV(j) = RV(j-1)[1-pac(j)^2] \quad , \; j=1,\ldots,NC \; , \qquad (4.2.10)$$

where RV(0)=1. It should be noted that $RV(j)=V_j$, and is regarded as the more stable method of computing the residual variances. The order determination function, CAT, is then calculated from the residual variances using the formula

$$CAT(k) = (1/N)\sum_{j=1}^{k} \frac{1-(j/N)}{RV(j)} - \frac{1-(k/N)}{RV(k)} \quad , \; k=1,\ldots,NC \quad . \qquad (4.2.11)$$

Next, we determine the AR order as the value of k which minimizes the function CAT(·), and denote it by k1. The coefficients, $a_{k1}(1),\ldots,a_{k1}(k1)$, of the AR(k1) model are then calculated by solving the Yule-Walker equations for order k1. Finally, the AR(k1) spectral density is calculated using the formula

$$f_{AR}(\omega_p) = RV(kl)\left|1+\sum_{j=1}^{kl} a_{kl}(j)\exp[2\pi ij\omega_p]\right|^{-2} \quad . \qquad (4.2.12)$$

4.3 The Graphical Results

The results for the first data set, the Wolfer Sunspots data, are presented in figures 3 through 5. Figure 3 shows the two periodograms which seem to represent the same spectral behavior. The smoothed periodograms in figure 4 show that most of the variability is concentrated in the frequency band 0 to 0.2 in which the two spectral estimates are very similar. By modeling the time series with an autoregressive scheme, the resulting spectral density estimates are, again, very close but not identical. The peak of the rank transform spectra is a little bit higher and occurs at a lower frequency. The general shape of the two spectral density estimates is, however, the same.

Figures 6 through 8 give the same results for the second data set, the Critical Radio Frequencies data. Here the results are even more convincing. The peaks are of the same height and occur at the same frequencies in all three spectral estimates, while the shape in general is almost identical.

We might mention here, as a general remark, that the results for the other time series we have checked which are not presented here were very similar to those here and sometimes even more amazing.

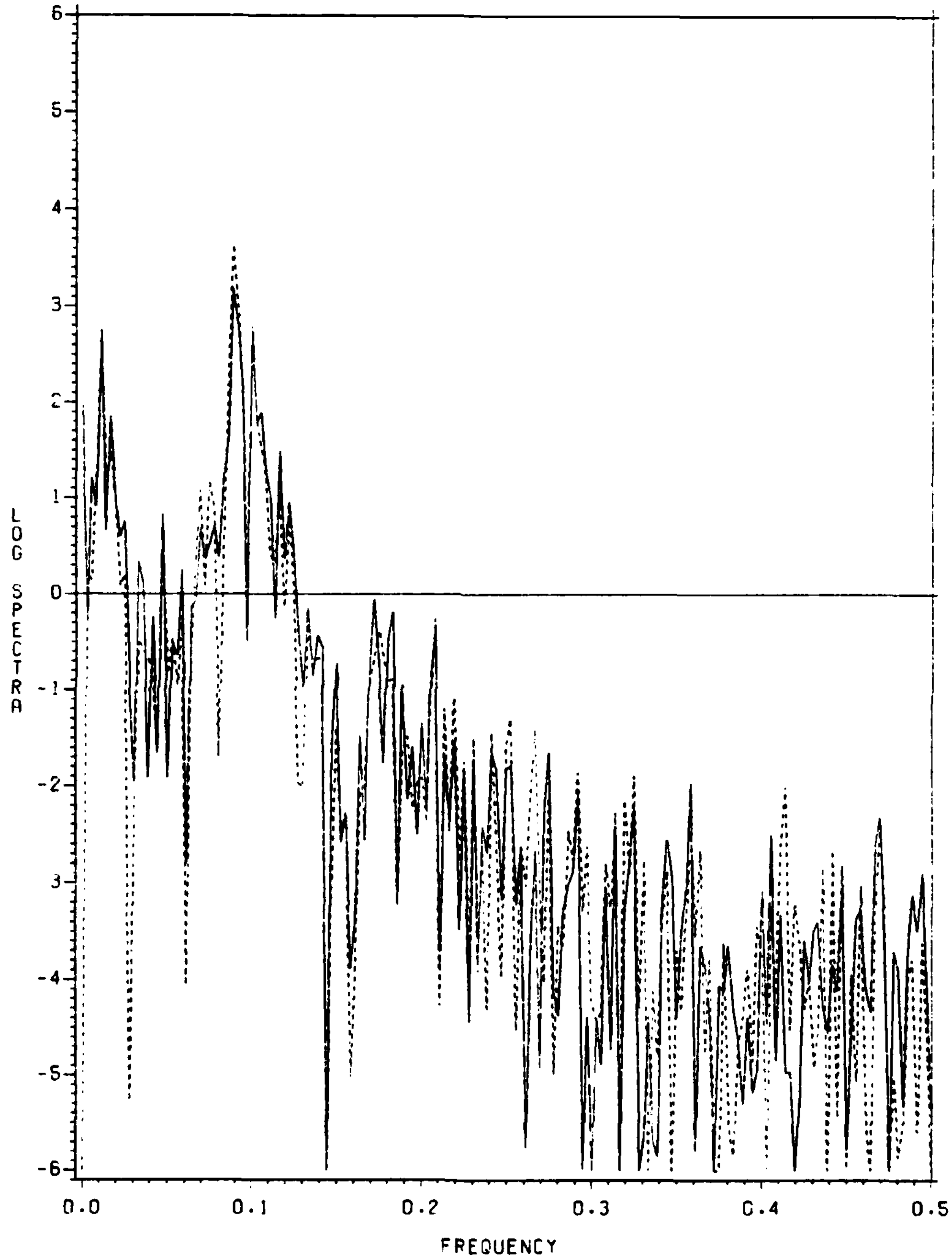


Figure 3. WOLFER SUNSPOT DATA - Raw periodogram. Solid line - original data, dashed line - rank transform data

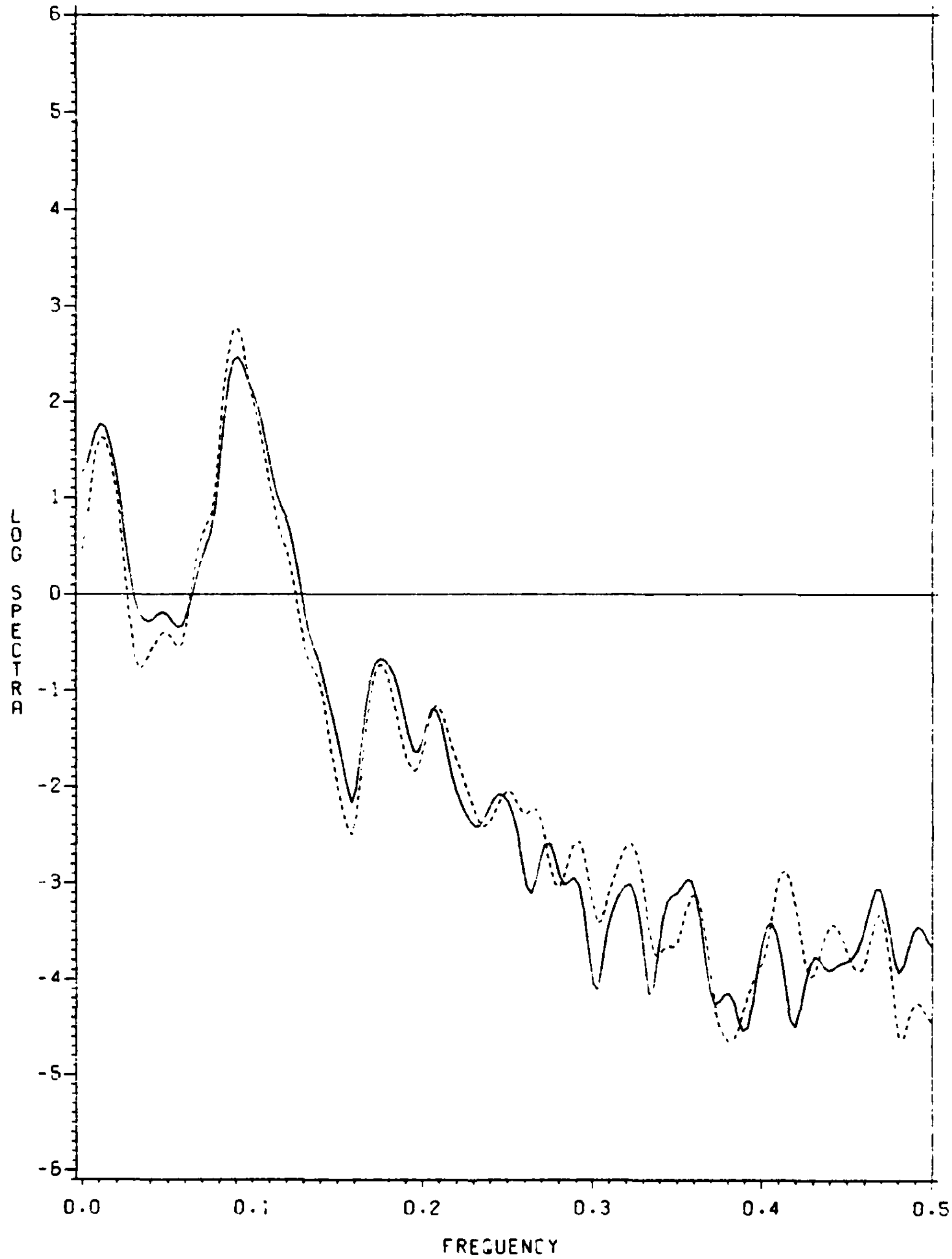


Figure 4. WOLFER SUNSPOT DATA - Smoothed periodogram. Solid line - original data, dashed line - rank transform data

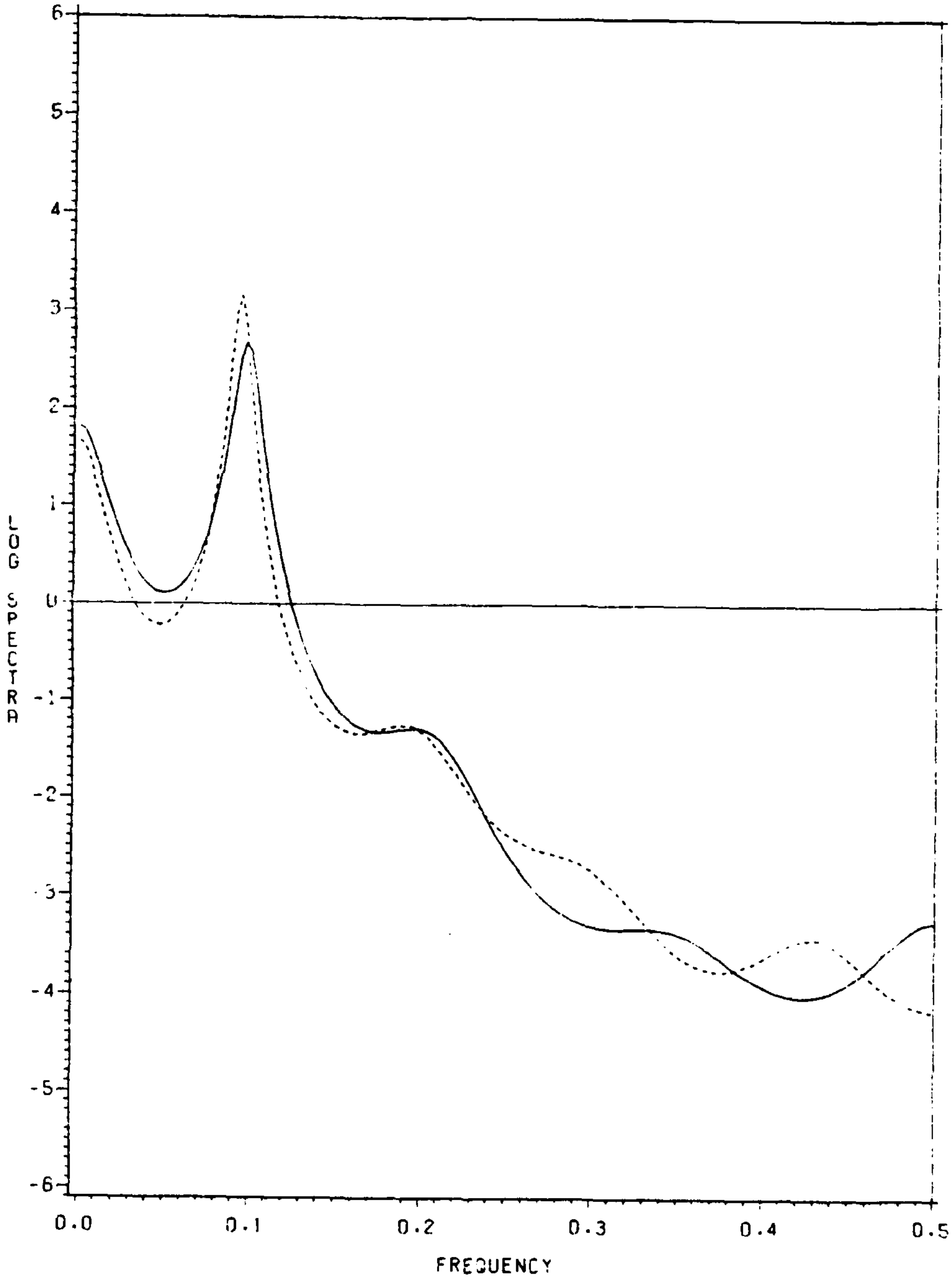


Figure 5. WOLFER SUNSPOT DATA - AR Spectral density. Solid line - original data, dashed line - rank transform data

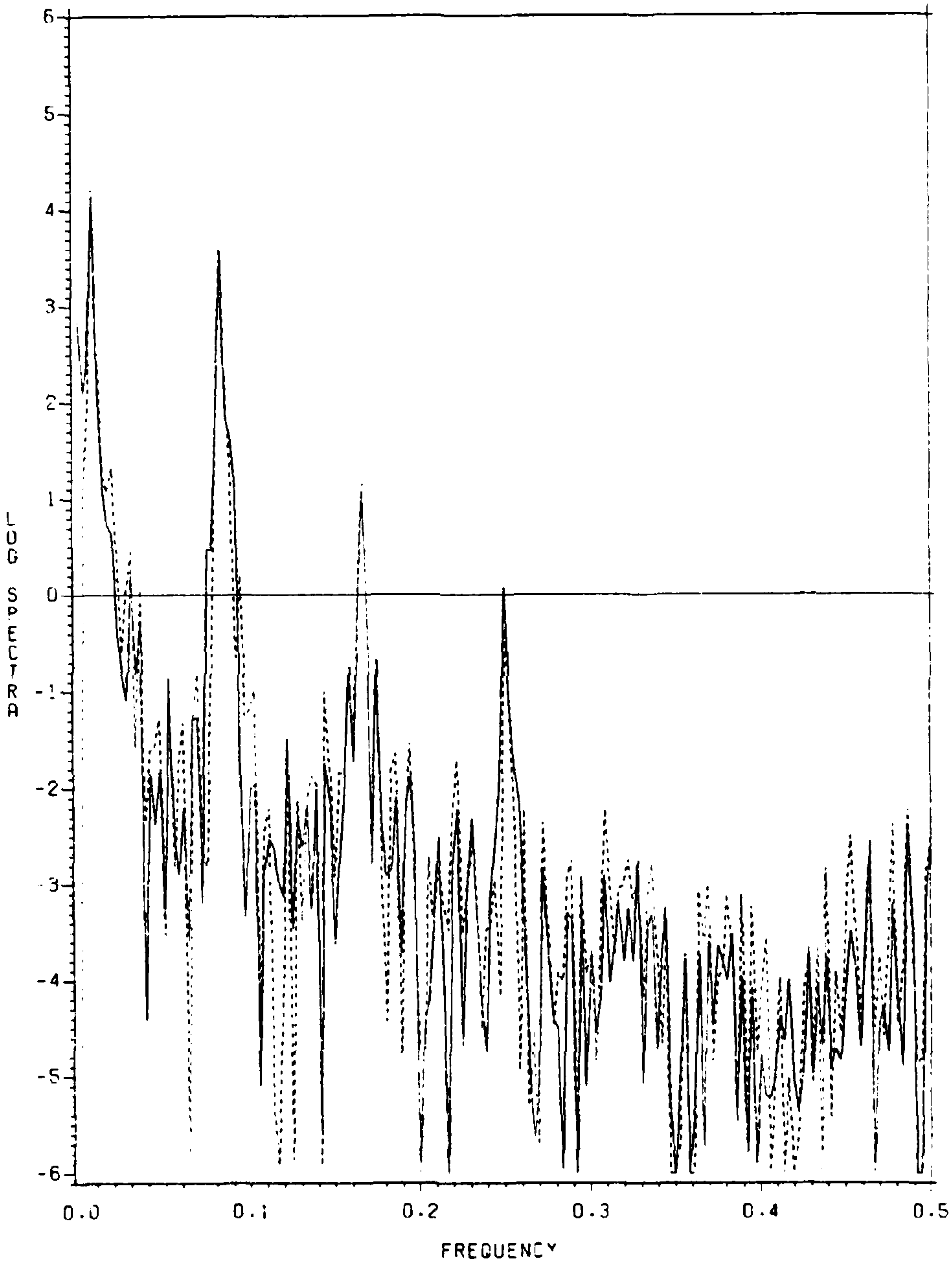


Figure 6. CRITICAL RADIO FREQUENCIES DATA - Raw periodogram. Solid line - original data, dashed line - rank transform data

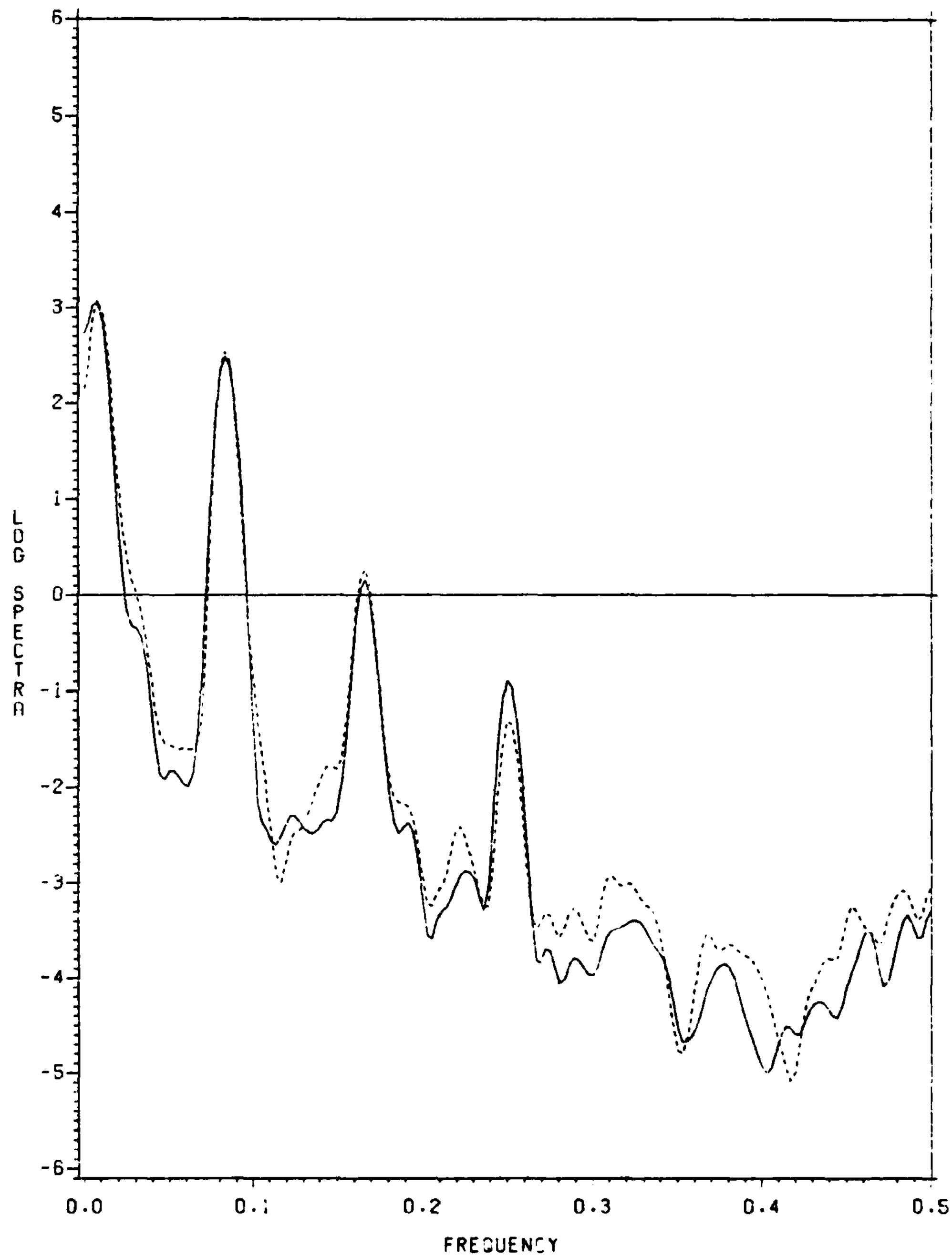


Figure 7. CRITICAL RADIO FREQUENCIES DATA - Smoothed periodogram. Solid line - original data, dashed line - rank transform data

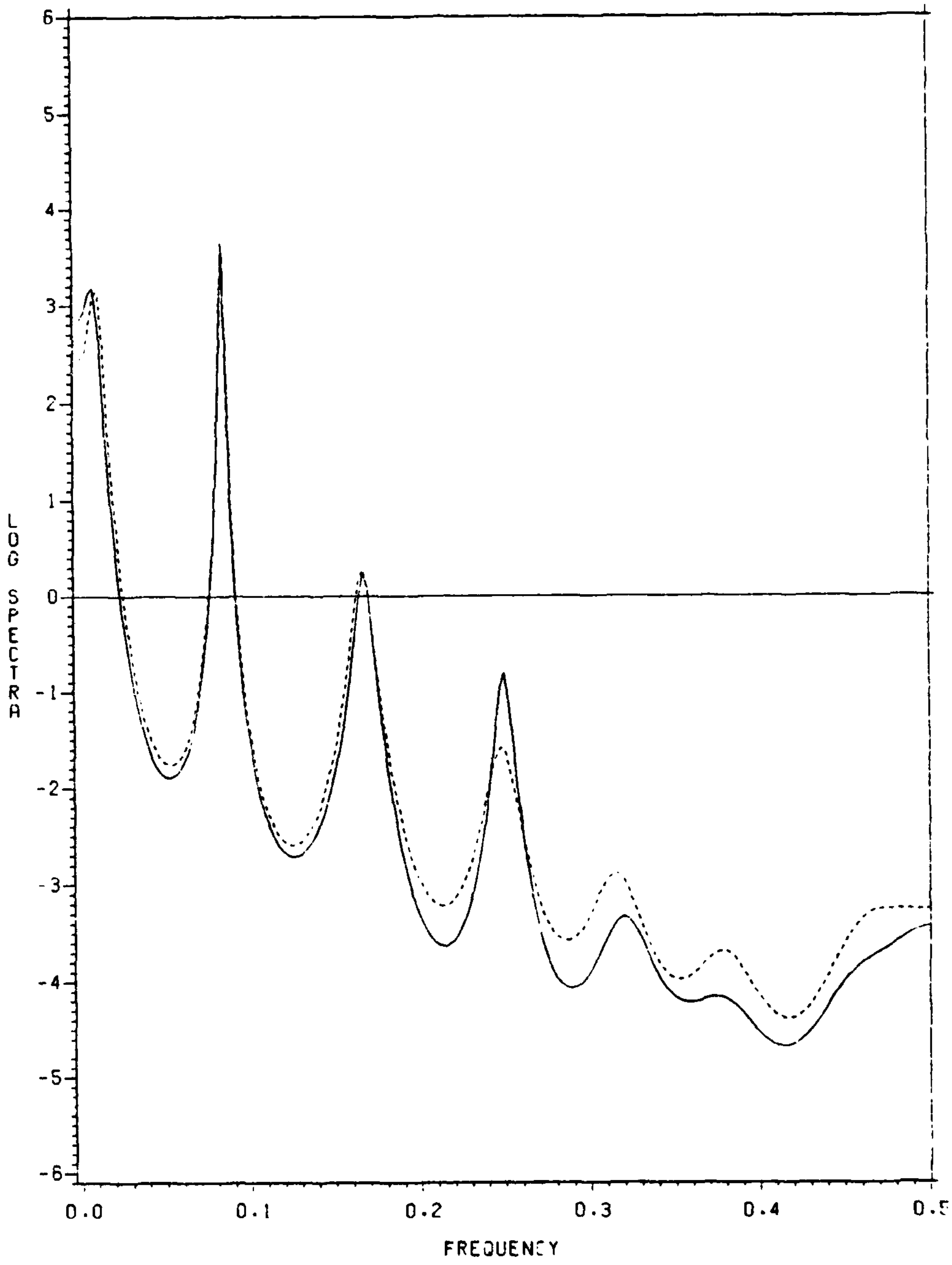


Figure 8. CRITICAL RADIO FREQUENCIES DATA - AR Spectral density. Solid line - original data, dashed line - rank transform data

CHAPTER V

CONCLUSION

5.1 Concluding Remarks

In this dissertation, the problem of applying nonparametric techniques to time series data is considered. The approach taken here is to extend existing theoretic results by interpreting and reformulating expressions for variance which are otherwise not easy to apply to statistical procedures. In particular, the asymptotic variance of linear rank statistics in the two sample problem, under dependence within the samples, is expressed in terms of the spectral densities of the corresponding rank transform time series. This result is then used to suggest estimators of the asymptotic variance by means of using existing time series analysis procedures to estimate the spectral densities of the rank transform time series. Also, this expression of the asymptotic variance enables one to better understand the conditions under which the asymptotic variance under dependence is greater or smaller than under independence.

The appearance of the rank transform spectrum in the expression for the asymptotic variance of linear rank statistics led us to examine the properties, in general, of the rank transform spectrum. Our impression, after examining several time series, is that the spectral density is approximately invariant under rank transformations. A report of this empirical investigation is given in chapter 4.

5.2 Problems for Further Study

The weak convergence of empirical processes formed from dependent data is an active area of research. Most of the published results in this area are for ϕ-mixing and strong mixing sequences. This presents a problem for applying those results in a time series situation since a time series obeying an autoregressive scheme does not, in general, satisfy the required mixing conditions. The notion of strong mixing Δ_s sequences, introduced by Gastwirth and Rubin (1975) solves the problem for some first order autoregressive schemes. It seems to be an open problem to prove a result for weak convergence of empirical processes that holds for a general ARMA(p,q) process.

Another problem for further research is to find other test statistics to which one can apply the approach used here to express the asymptotic variance in terms of the spectral density of some related time series.

Our impression from working with small samples and with long memory time series is that one can gain valuable information by applying asymptotic techniques even when the data doesn't seem to obey the required assumptions. It is hence desirable that theorems be complemented by data analytic diagnostics usable by applied statisticians.

In this work we have done a limited empirical investigation of the properties of rank transform spectrum. The theoretical properties, however, of the rank transform spectrum are still an open research problem.

REFERENCES

Babu, G. J. and Singh, K. (1978), "On Deviations Between Empirical and Quantile Processes for Mixing Random Variables," *Journal of Multivariate Analysis*, 8, 532-549.

Bahadur, R. R. (1966), "A Note on Quantile in Large Samples," *Annals of Mathematical Statistics*, 37, 577-580.

Billingsley, P. (1968), *Convergence of Probability Measures*, New York: John Wiley.

Blomqvist, N. (1950), "On a Measure of Dependence Between Two Random Variables," *Annals of Mathematical Statistics*, 21, 593-600.

Csörgő, M. and Révész, P. (1981), *Strong Approximations in Probability and Statistics*, New York: Academic Press.

Falk, M. and Kohne, W. (1984), "A Robustification of the Sign Test Under Mixing Conditions," *Annals of Statistics*, 12, 716-729.

Fears, T. R. and Mehra, K. L. (1974), "Weak Convergence of a Two-Sample Empirical Process and a Chernoff-Savage Theorem for ϕ-Mixing Sequences," *Annals of Statistics*, 2, 586-596.

Gastwirth, J. L. and Rubin H. (1975), "The Asymptotic Distribution Theory of the Empiric CDF for Mixing Processes," *Annals of Statistics*, 3, 809-824.

Gleser, L. J. and Moore, D. S. (1983), "The Effect of Dependence on Chi-Squared and Empiric Distribution Tests of Fit," *Annals of Statistics*, 11, 1100-1108.

Kiefer, J. (1970), "Deviations Between the Sample Quantile Process and the Sample D.F.," *Nonparametric Techniques in Statistical Inference*, M. L. Puri, ed., London: Cambridge Univ. Press.

Mehra, K. L. and Rao, M. S. (1975a), "On Functions of Order Statistics for Mixing Processes," *Annals of Statistics*, 3, 874-883.

__________ (1975b), "Weak Convergence of Generalized Empirical Processes Relative to d_Q Under Strong Mixing," *Annals of Probability* 3, 979-991.

Owen, D. B. (1980), "A Table of Normal Integrals," *Communications in Statistics, Simulation and Computation*, B9(4), 389-419.

Parzen, E. (1979), "Nonparametric Statistical Data Modeling," *Journal of the American Statistical Association*, 74, 105-131.

_________ (1983), "FUN.STAT Quantile Approach to Two Sample Statistical Data Analysis," Technical Report No. A-21, Texas A&M University, Department of Statistics.

Pyke, R. and Shorack, G. R. (1968), "Weak Convergence of a Two-Sample Empirical Process and a New Approach to Chernoff-Savage Theorems," *Annals of Mathematical Statistics*, 39, 755-771.

Sen, P. K. (1971), "A Note on Weak Convergence of Empirical Processes for Sequences of ϕ-Mixing Random Variables," *Annals of Mathematical Statistics*, 42, 2131-2133.

Serfling, R. J. (1968), "The Wilcoxon Two-Sample Statistic on Strongly Mixing Sequences," *Annals of Mathematical Statistics*, 39, 1202-1209.

_________ (1980), *Approximation Theorems of Mathematical Statistics*, New York: John Wiley.

Yokoyama, R. (1973), "Weak Convergence of Empirical Processes for Strong Mixing Sequences of Random Variables," *Science Reports. Tokyo Kyoiku Daigaku. Section A*, 12, 36-39.

_________ (1980), "A Note on Weak Convergence of Empirical Processes for ϕ-Mixing Random Variables," *Yokohama Mathematical Journal*, 28, 37-40.

Yoshihara, K. I. (1978), "Note on Multidimensional Empirical Processes for ϕ-Mixing Random Vectors," *Journal of Multivariate Analysis*, 8, 584-588.

END

FILMED

1-86

DTIC